Evaluating the Norfolk Physical Education Program with Respect to Fitness-Level Improvement

Robert Wayne Baker, B.A., B.Ed.

Department of Graduate and Undergraduate Studies in Education

Submitted in partial fulfillment of
the requirements for the degree of
Master of Education
Faculty of Education
Brock University
St. Catharines, Ontario
September, 1990

Abstract

The grades seven and eight physical education program of the Norfolk Board of Education was evaluated with respect to fitness-level improvement, an objective of the Ministry of Education for the province of Ontario. The Canada Fitness Award battery of fitness tests was used to measure fitness levels.

It was established that in September the students were unfit, and in May they were fit. This indicated that the Norfolk physical education program was effective, with respect to the criterion used for this research.

In addition, it was discovered that fitness-level improvement was significantly related to certain variables: teacher qualifications, teaching experience, school, and participation in extracurricular physical activity.

Considering the results of the research, it was recommended that the Norfolk Board of Education hire young, qualified physical education teachers; create the position of Physical Education Consultant; and strive to create equitable resources for physical education instruction, in order that the school to which a student belongs no longer will be a determinant of fitness improvement.

Acknowledgements

Completion of this thesis was a formidable task, made easier by the contributions of several persons.

First of all, I would like to thank Dr. John Kearns for his ongoing support. Not only did he provide expertise within the field of physical education, but he was always available with a vote of confidence.

Direction in the area of statistics was provided by Dr. Kris Kirkwood. It was a comfort to have such a talented resource at my disposal. In addition, Kris acted as second reader for my research.

An old friend, Dr. Wally Poole, also acted as a reader for this research. His opinion is one which I value greatly.

Mr. Tony Sylvestre was another valuable resource. Through countless hours at the computer terminal, Tony forced me to rethink and refine my original research hypotheses.

Finally, and most importantly, I would like to dedicate this research to Mary Anne, Michael, and Amanda Baker, who did without a husband and father far more than they deserved.

TABLE OF CONTENTS

LIST OF TABLES

Table		Page

Table		Page

CHAPTER I

THE PROBLEM

Purpose of the Study

During his tenure as Minister of Education for the province of Ontario, Chris Ward has asserted that one goal of education should be to "develop physical fitness and good health" (Ontario Schools: Intermediate and Senior Divisions, Grades 7-12/O.A.C.'s, 1989, p. 3). With this consideration in mind, it is apparent that physical education programs must be concerned to a major degree with individual student physical fitness levels. The concern of this research is identical.

The purpose of this research, therefore, is to evaluate the effectiveness of the grades seven and eight physical education program currently in place within the Norfolk Board of Education by examining student fitness levels.

Scope of the Study

This study is an evaluation of the physical education program of Norfolk, using fitness-level improvement as the criterion of effectiveness. The study period is one school year, involving all grade seven and eight students who participate in physical education. To avoid potential health risks, no student who, for health reasons, does not regularly participate in physical education classes, is allowed to participate in the study.

Importance of the Study

Student behaviours can change as a result of physical education instruction (Loughery, 1987). The most pertinent of these behavioural changes is "physical changes".

For the purpose of this research, therefore, program effectiveness is considered in terms of the degree to which physical changes (in strength, power, speed, agility, muscular endurance, and cardio-respiratory endurance) occur.

Definition of Terms

There are several terms to be defined for the purpose of this research.

1. Physical Fitness-Physical fitness is the ability to enjoy, and function actively within, one's life without undue fatigue.

   For this research an appropriate level of fitness would correspond to the 50th percentile of the Canada Fitness Award scores, hence, an average fitness level.

2. Program-Program refers to the cumulative physical education instruction received by Norfolk Board grade seven and eight students. In most cases the program offered is a reflection of the Ministry of Education (1978) curriculum guidelines, which suggest seven categories of activities to be included in the Intermediate physical education program: physical

fitness activities, team sports, individual and dual activities, gymnastic activities, dance, outdoor education, and aquatics.

3. Program Effectiveness-Program effectiveness will be determined by the degree to which the physical education program improves the mean fitness level of Norfolk grade seven and eight students. An effective program will significantly improve the mean fitness level of the Norfolk students.

4. Mean Fitness Score-A mean fitness score reflects an "average" fitness test performance. A mean fitness score was determined for each of the six subtests within the Canada Fitness Award; hence, there was a mean "push up" score, a mean "shuttle run" score, a mean "50-metre run" score, a mean "partial curl-up" score, a mean "endurance run" score, and a mean "standing long jump" score.

A mean fitness score was also calculated for the combined test components. This was accomplished by assigning a point score for each of the six Canada Fitness Awards tests. If the Excellence level was achieved, four points were awarded. Three points were earned for the Gold level; two points for the Silver level; one point for the Bronze level; and zero points for scoring below the Bronze level. The maximum score for any student was twenty-four, while the minimum was zero. The mean composite score was the average of these scores; there was no weighting of scores.

5. Physical Education-"Physical education is that portion of the educative process which utilizes physical activity as a primary means for influencing the psychological, intellectual, and social, as well as the physical development of the individual to effectively meet and adjust to the demands of a changing society" (Van Holst, 1983, p. vi).

6. Canada Fitness Award (CFA)-The Canada Fitness Award is a "fitness incentive program for Canadians from six to seventeen years of age" (Fitness Canada, 1985, p.3).

The Canada Fitness Award consists of six individual tests: push ups, partial curl-ups, shuttle run, standing long jump, 50-metre run, and an endurance run (800, 1600, or 2400 metres, depending on age). Performance scores are recorded for each of the subtests.

An aggregate fitness score is also developed. The raw performance score for each subtest is translated into a point value of 0, 1, 2, 3, or 4, based upon the CFA level achieved (0 for not achieving the Bronze level and 4 for achieving the Excellence level). The six point scores (there are six subtests) for each student are totalled to render the aggregate fitness score, a score indicative of overall fitness. This aggregate score is the criterion by which individual quartile rankings are assigned; the range of Norfolk scores is divided into four equal portions and rankings assigned accordingly.

7. <u>Quartiles</u>-The lower quartile divides the lowest 25% of values from the upper 75%, while the upper quartile separates the highest 25% of values from the lower 75%. The middle quartile is the median of the distribution of values. By establishing these three quartiles the distribution is effectively divided into four equal (in terms of total number) portions of 25%.

   In this research each student was assigned in September to one of these four groups. For ease of identification the lowest 25% of scores was referred to as "Level 4", the next 25% "Level 3", the following 25% "Level 2", and the top 25% "Level 1". This ranking system was useful in discovering if initial fitness standing was a determinant of fitness-level improvement.

8. <u>Strength</u>-Strength can be described as the force exerted during a maximal effort (Fitness Canada, 1985).

9. <u>Speed</u>-Speed refers to the ability to move from one place to another in the shortest possible time (Gabbard, LeBlanc, & Lowy, 1987).

10. <u>Muscular Endurance</u>-"Muscular endurance involves the ability of a group of muscles to repeat continuously a performance requiring a relatively high level of muscular force or to maintain a position against a force that is counteracting it" (Greenberg & Pargman, 1986, p. 77).

11. <u>Agility</u>-"Agility refers to the controlled ability to change position and direction rapidly and accurately" (Fitness Canada, 1985).

12. <u>Cardiorespiratory Endurance</u>-Cardiorespiratory endurance (or cardiovascular endurance or aerobic endurance) is "...the ability of the heart, lungs, and vascular system to function efficiently for an extended period of time" (Gabbard, LeBlanc, & Lowy, 1987, p. 52).

13. <u>Health-Related Fitness</u>-Health-related fitness concerns itself with the ability of the individual to function in everyday life. The components of health-related fitness are strength, cardiorespiratory function, muscular endurance, and flexibility (Pifer, 1987).

These factors are believed, to a certain extent, to protect against heart disease, obesity, and musculoskeletal disorders (Pate, 1983).

14. Performance-Related Fitness-Performance-related fitness is concerned with the level of functioning that enables the individual to participate in sports.

15. Physical Activity-Physical activity is "body movement that is produced by the muscles, and results in a significant expenditure of energy" (Shephard, 1986, p. 4).

16. Body Composition-Body composition is the relative percentage of fat and fat-free body mass (Blair, Falls, & Pate, 1983).

17. Flexibility-Flexibility refers to the range of motion through which the body can move.

Background of the Problem

There are several unresolved issues related to physical education, two of which relate to this research.

While research has suggested the benefits of daily physical education (Grace, 1987; Rippe, 1987; Ross & Pate, 1987; Shephard, 1986), others, including professional educators and laymen alike, have questioned the value of physical education with respect to improving physical fitness (Beauchamp, 1980).

In addition, physical educators need to know whether their instruction is of benefit to their students in terms of fitness improvement.

These issues can be resolved only after it has been determined whether or not the existing physical education program is effective.

Evidence of an effective program would provide fuel for the movement toward daily physical education, answer skeptics who would surrender physical education time to other subjects, and provide physical educators with positive feedback concerning their efforts.

Evidence of an ineffective program would suggest a different course of action: curriculum and teacher review.

Research Objectives

There are several questions to be answered in the course of this study.

Is the Norfolk grade seven and eight physical education program effective? Are Norfolk elementary physical education teachers improving the fitness levels of their students? Might changes be necessary in such areas as curriculum material or personnel assignment? These questions provide the impetus for this research.

The specific questions within this research are as follows:

1) Were grade seven and eight students in Norfolk fit (according to Canada Fitness Award standards), both in September and May?

2) Was there a significant improvement in fitness levels over the course of the school year?

3) If there was a significant improvement, was it related to the variables included in this research: gender, age, grade, teacher qualifications, instruction time per cycle, nature of the class, initial fitness level, or participation in physical activity outside of the regular physical education class?

Conceptual Assumptions

1. From previous research it could be expected that fitness levels would prove to be higher on the post-test than on the pre-test (Committee for the Development of Sport, 1983; Pifer, 1987). In effect, then, it is assumed that physical education instruction has a positive effect on fitness levels. According to Bozzo (1983), this is a reasonable assumption.

2. It is assumed that the concept of fitness is measurable. There is an abundance of fitness tests to substantiate this assumption.

3. It is assumed that fitness improvement can be measured. If fitness can be defined and measured, the measurement of improvement becomes a numerical exercise. Mean scores reflect the "average" Norfolk performances on each test. September and May mean scores are compared using *t*-tests in order to determine the significance of fitness-level change.

4. The ability to measure program effectiveness is assumed. For the purpose of this research an effective program demonstrates a significant improvement in the fitness scores of students.

5. It is assumed that fitness-level improvement constitutes program effectiveness. The Ministry of Education (1978) cited fitness improvement as a goal of physical education programs. Loughery (1987) concurred. He stated that physical changes, arguably reflected in fitness improvement, could be considered in determining an effective physical education program.

Rationale for the Study

The rationale for this research focusses upon the need for ongoing evaluation of program effectiveness. By examining similar research, it is evident what results and characteristics previous researchers considered indicative of successful, or effective, programs. As will be indicated in Chapter II, much of the previous research into program effectiveness employed a less rigorous definition of effectiveness than does this research.

Delineation of the Research Problem

Fitness-level change, the criterion for measuring program effectiveness, is examined for relationship with several factors: gender, age, grade, teacher qualifications, nature of the physical education class, instruction time per cycle, participation in extracurricular physical activity, and initial fitness level. These variables were chosen based largely upon past research.

Gender

The choice of gender as an independent variable addresses the question: Does one gender benefit more from the physical education instruction offered in Norfolk than the other? There is much research to suggest that males perform better on fitness tests (Alexander et al., 1985; Greenberg & Pargman, 1986; Raithel, 1987), but can they also be expected to improve more?

Age

The choice of age as an independent variable addresses the question: Is physical education instruction more beneficial to certain ages? This question is important because of the age range of grade seven and eight students: the onset of puberty is often experienced during these years and puberty has shown itself to be a factor in fitness testing (Bozzo, 1983).

Grade level

Grade level is of interest as an independent variable because it could be expected that, all other variables being equal, grade eights should score higher on fitness tests since they have had an extra year of physical education instruction. But can they be expected to improve more?

Instruction time

The choice of instruction time per six-day cycle as an independent variable approaches the issue of daily physical education instruction. Past research (Cooney, 1987; Grace, 1987; Murphy, 1987a) suggests that the fitness levels of students participating in daily physical education are higher than for other students, but is it realistic to expect greater improvement also?

In addition, while there is evidence to support the benefits of daily physical education, this research will explore the value of increased, but less than daily, physical education instruction time.

Extracurricular physical activity

Participation in extracurricular physical activity indicates physical activity levels higher than those not involved in activities outside the school (Caspersen, 1987; Ross & Pate, 1987; Smith & Gilligan, 1987). This should result in higher fitness levels, but can greater improvement also be expected?

Teacher qualifications

The matter of teacher qualifications is concerned with certain demographic characteristics, most notably whether or not the teacher has a physical education degree, a physical education specialist's certificate, or any other physical education training. Based upon past research, such as that of Ross and Pate (1987), Grace (1987) has suggested that teachers in charge of physical education classes should be appropriately trained. The inference is that students are more fit when the physical education teacher is trained. Can these students also be expected to improve more?

Nature of the physical education class

The nature of the physical education class addresses the matter of segregated versus co-educational classes. The Ontario Medical Association (1987) has suggested that co-educational classes are not as effective as segregated classes in improving the fitness levels of all students involved. According to this stance, students who participate in segregated physical education classes should score higher on fitness tests. But, should they also improve more?

Initial fitness level

Initial fitness level of the student relates to the fitness level of the student before receiving physical education instruction in 1988-89. Past research (Shephard, 1986) has suggested that much more effort is required to improve higher levels of fitness than is necessary to raise lower levels of fitness. For this research the aggregate fitness scores assigned to students after the September testing will be examined for relationship with fitness-level improvement.

Statement of Hypotheses

For statistical purposes, the research objectives were restated as null hypotheses (H_0#1-3):

H_0#1 There will be no significant difference between the average (mean) Norfolk fitness scores and the average (50th percentile) Canada Fitness Award scores.

H_0#2 There will be no significant difference between pre- and post-treatment mean scores for Norfolk students on any of the tests.

H_0#3 There will be no significant positive relationship between fitness-level improvement and any of the variables tested.

CHAPTER II

REVIEW OF RELATED LITERATURE

Chapter Overview

Chapter II attempts to narrow the focus of the research from the broadest discussion of the concerns within the field of physical education to the specific purpose of this research.

The chapter consists of a review of literature and research relating to:

a) the background to the physical education/fitness dilemma

b) the purposes of evaluating educational programs

c) methods of evaluating physical education programs

d) defining physical fitness

e) the delineation of fitness components

f) tests that have been substantiated as measuring these fitness components

g) test batteries which exist

h) the test battery for this research

i) determinants of fitness performance/improvement

j) the contextual framework for this study

History of Fitness Research

Because of the current nature of the issue, much of the research into fitness measurement and its related variables is found in journals.

The American Alliance of Health, Physical Education, Recreation, and Dance (AAHPERD), interested in fitness measurement since the 1950's, established the foundation for current fitness research by developing the AAHPER Youth Fitness Test Manual (1965), the Youth Fitness Test (1976), the AAHPERD Health-Related Physical Fitness Test (1980), and the "Physical Best" Fitness Test (1989).

The Canadian Association for Health, Physical Education, and Recreation (CAHPER) took its lead from AAHPER, developing the CAHPER Fitness-Performance Test (1966; 1980) and the Canada Fitness Award (1970; 1979).

The Journal of Physical Education, Recreation, and Dance (JOPERD) has been active in fitness research, publishing the findings of Falls (1980) and Pate and Ross (1987), which re-defined current thought concerning fitness and health.

The Physician and Sportsmedicine, the Journal of Sportsmedicine and Physical Fitness, and the Physical Educator have also addressed concerns of fitness testing.

Several books provided essential background for this research. In particular, <u>Concepts in Physical Education</u> (Corbin, Dowell, Lindsey, & Tolson, 1981), <u>Measurement and Evaluation in Physical Education, Fitness, and Sports</u> (Bosco & Gustafson, 1983), <u>Physical Fitness: A Wellness Approach</u> (Greenberg & Pargman, 1986), <u>Physical Education for Children: Building the Foundation</u> (Gabbard, LeBlanc, & Lowy, 1987), and <u>Bases of Fitness</u> (Fox, Kirby, & Fox, 1987) provided the author with the current concepts and definitions of fitness.

Several unpublished studies were of assistance also. These tended to place the theory into a more practical application. Of particular importance were: Understanding Fitness: A Primer for School Program Planners (Beauchamp, 1980); Evaluation of Programs for the Promotion of Physical Fitness and Exercise (Bozzo, 1983); A Collection of the Conclusions of the European Seminars Organized by the Committee for the Development of Sports (CDDS, 1983); Factors Influencing Physical Fitness Evaluation (Harrer, 1983); and A Tool for Assessing and Designing a K-12 Physical Education Program (Iowa State Department of Public Instruction, 1985).

Background to the Physical Education Dilemma

It could be asked why the physical education program of Norfolk warrants attention, or why changes in physical education curriculum or personnel might be in order.

Over the past several years physical educators have become concerned about the health and fitness levels of children and adolescents. The American Medical Association (AMA) has stated that 40% of school-aged children are overweight (Gabbard, LeBlanc, & Lowy, 1987): 10-20% over the recommended weight relative to the height of the individual. Giel (1988), Hoerr (1987), and Pate (1983) contend that 20% of American children are obese: more than 20% overweight. According to the National Youth Fitness Study of the United States (West, Beveridge, & Workman, 1987), figures such as these illustrate that American children are less fit than in previous studies.

The situation in Canada is less clear. In Bailey (1973), Goode (1976), and again in Beauchamp (1980), it has been claimed that school children are unfit. The Ontario Medical Association (OMA) has also noted an increasing incidence of obesity in children (Grace, 1987). However, Gauthier, Massicotte, Hermiston, and MacNab (1983) compared the work capacity of Canadian children from 1968 to 1983, and concluded that there had been a significant improvement in physical fitness during that time period. In addition, Canada Fitness Award standards have risen over the twenty-year existence of the test, suggesting improved performances.

This dichotomy can be partially explained through further examination of the findings of Gauthier et al (1983). It was discovered that the most significant improvement in fitness levels occurred with 13-to-17 year old girls. It was suggested that girls improved because of increased participation in athletic programs since 1968. In addition, there was not a significant fitness-level improvement for all boys. These facts might indicate that the lower end of the fitness continuum has improved but that these improved fitness levels are still substandard. Also, during the six years since the research by Gauthier, children have spent increasing amounts of time with computers and other passive forms of entertainment (Raithel, 1987). It could simply be that fitness levels have fallen since 1983.

To further refute the argument of improving fitness levels, the research of Dahlgren (1982), concerning the Canada Fitness Award, is useful. She noted that, although scores for three of the six tests "generally improved", this did not necessarily indicate improved fitness. Possible explanations could be familiarity with the test or practice effect.

This difference of opinion supports the contention of Giel (1988), who stated that it is difficult to compare the fitness levels of children over time because of the changing definitions of fitness and the changing philosophies about fitness testing.

Physical activity and fitness

It has been well established that overweight conditions are associated with low levels of physical activity (Cooney, 1987), and the Ontario Medical Association (OMA) stated that children are up to 40% less active than their counterparts of thirty years ago (Cooney, 1987). These reduced activity levels manifest themselves in concomitantly lower fitness levels. We could blame the sedentary lifestyle of modern society, demanding less physical exertion on the job and in the home. However, many people have placed blame on physical education programs for allowing fitness levels to degenerate.

Bailey (1973) considered that "for the ordinary Canadian child physical fitness...seems to be a decreasing function of age from the time we put him behind a desk in our schools," (as quoted in Beauchamp, 1980, p. 1). As school years pass, the situation seems to worsen. The Saskatchewan Growth Study found that children show a significant decline in physical fitness beyond the age of twelve (Cooney, 1987).

With these condemnations in mind, the onus has fallen upon physical educators to defend themselves against critics. As stated in Iowa State Department of Public Instruction (1985), "...the role of physical education in the public school setting has been...questioned and challenged by the public..." (p. 5). In Canada, Goode (1976) and Beauchamp (1980) have also questioned the apparent value and effect of physical education in terms of fitness development.

It is difficult to be convinced that the education system can be held solely accountable for the physical condition of Canadian youth. Educators cannot entirely alter values which have been ingrained at home. Decreased activity levels, as well as ever-increasing hours of passive entertainment (Ross & Pate, 1987) -video games and television, especially-, in addition to overeating, reflect values inculcated into young people. As with most other value systems, schools can do little to change them completely. If, however, the education system accepts some responsibility for changing the situation, and it must, then physical educators must not be forced to act in isolation.

Daryl Siedentop feels that too few people care about high school physical education (Pifer, 1987). Dan Cooney, formerly of the Canadian Association for Health, Physical Education, and Recreation (CAHPER), might argue that too few people care about physical education at any level. Cooney (1987) cited a declining interest in physical education even among educators: "fewer central office physical education consultants, 'greying' of physical education teachers and other teachers, lack of interest in physical education programs by Ministry of Education officials, and optional physical education courses in high school" (p. 15).

Most physical educators could cite personal examples of this feeling of disconcern. Within the Norfolk Board, the setting for this research, there is not, nor has there ever been, a physical education consultant. The scheduled Ministry of Education curriculum guidelines for physical education (Spring, 1989) will be the first of their kind since 1978. Considering the changes that have occurred within the field of physical education since 1978, eleven years is a long time between curriculum assessments. Finally, students may opt out of physical education in the Intermediate division at the discretion of their parents (Ontario Ministry of Education, 1978). With the exception of religious education, there is no other subject with which students have that option.

Physical educators feel that they could partially remedy the fitness problem, but not without the assistance of administrative timetabling. At the Canadian Summit on Fitness, 1986, the goal was unanimously adopted for the entire school system: "to implement...daily physical education and activity programs as part of a core curriculum from preschool to post-secondary institutions" (OPHEA, 1988, p. 9). Several studies have indicated the benefits of daily physical education (Martens, 1982; Murphy, 1987a; Pritchard, 1987) and, although movement in that direction has been slow, endorsement of the concept continues (CAHPER, 1986; Cuniff, 1985; Ontario Medical Association, 1987). Fitness Canada has gone so far as to provide funding for CAHPER to promote Quality, Daily Physical Education (OPHEA, 1988).

The move to daily physical education would be drastic when one considers that at present far less than daily instruction time is allotted. Bozzo (1983) stated that only half of American elementary students get physical education twice a week and Beauchamp (1980) asserted that the majority of Canadian elementary schools schedule only two classes of physical education per week. According to the Ontario Medical Association (1987), two classes of physical education per week seems insufficient to improve fitness.

At present only about one-third of American youth, aged ten to seventeen, participate in daily school physical education programs and the proportion is declining steadily (Cuniff, 1985; McGinnis, 1988), again reflecting time-demand conflicts. To begin to improve physical fitness levels, Cuniff (1985) feels that over 60% of all students should be receiving daily physical education instruction by 1990. Cooney (1987) also thinks that additional time should be given to physical education instruction, not solely for educational reasons, but also as a reflection of societal views. The fitness boom, which is rampant among adults (Lashuk, 1984), is not reflected in the allotment of instruction time. In British Columbia, Alberta, Saskatchewan, Ontario, New Brunswick, and Nova Scotia, physical education instruction time has not increased concomitantly in recent years (Cooney, 1987).

The physical education dilemma has been established. Physical educators are concerned that children and adolescents are not developing fitness levels that can help them maintain good health (Bozzo, 1983); in order to correct this situation, physical educators are asking for more instruction time. Skeptics, on the other hand, are questioning the contribution of physical education to the physical well-being of students. As a result, the amount of physical education instruction time is, at best, being maintained or, at worst, being eroded. Regardless of how the amount of instruction time might change, a concern exists among physical educators over the poor performance by children and adolescents in certain fitness components, especially cardiovascular endurance (Murphy, 1986).

The Purpose of Evaluating Educational Programs

The ultimate purpose of evaluating any educational program is considered by Loughery (1987) to be the assessment of a program's impact or effectiveness. In this research, the impact of physical education instruction on student fitness levels is the criterion of evaluation. Lazarus (1982) enumerated five secondary goals, as identified by the Evaluation Research Council, for evaluating any program (as cited in Loughery, 1987).

Front End Analysis determines the need for a new program and subsequently decides upon the "necessary levels of support" if the new program is, in fact, initiated. Obviously, if a program already exists, front end analysis is impractical. However, this research could be considered Front End Analysis in the investigation of a potential new program: daily physical education. Considerations, or "levels of support", would include: availability of instruction time and qualified personnel for presentation of the adapted program.

Evaluability Assessment focusses upon the essential need of a new program to be evaluated. Once a program is instituted, how can we decide if it has accomplished what it was intended to do? Gabbard, LeBlanc, and Lowy (1987) also identified this purpose for program evaluation, asking the question: "How well did the program meet the established objectives?"

Formative Evaluation is concerned with assessing the ongoing process of a program. Does our evaluation indicate that the program is moving positively toward the accepted goal?

Evaluation of an Evaluation involves the collection of new data, or reanalysis of past data, about a program in order to determine the validity and reliability of the evaluative process.

The assessment of a program's effectiveness must be directed toward the "product" of our educational system: the student. How successful is the process of instruction for the student? The Iowa State Department of Public Instruction (1985) also acknowledged the need for program assessment to focus upon the student when they were concerned with "...the status of the program and how well the status of program offerings is meeting the student needs that have been targeted..." (p. 17). By so doing, Lazarus (1982) and Jewett and Bain (1985) suggested that the instructional process can be improved. The author hopes that this research adds to the instructional process at work within Norfolk.

One purpose of this research is to provide feedback to Norfolk physical educators concerning the success of their program. In other words, this study is monitoring the physical education situation in Norfolk. Program monitoring was defined by Loughery (1987) as "the periodic examination of programs that might focus on issues of program compliance or the collection of basic descriptive data that might assist program management" (p. 64). Personalizing that view, this research will provide descriptive data in September and May in order to ascertain if the Norfolk physical education program is operating in compliance with Board and Ministry of Education mandates, namely, to develop physical fitness (Ontario Ministry of Education, 1978). For any program evaluation to be effective, then, it must be clear exactly what the program was expected to do for its clients (Loughery, 1987).

There are several reasons why determining the effectiveness of physical education programs is desirable. They relate to the unresolved issues which exist within the field of physical education.

First, recent research has suggested the benefits of daily physical education (CAHPER, 1986; Cuniff, 1985; Ontario Medical Association, 1987). In most instances the move to daily physical education would require a substantial reallocation of available instruction time. Before such a move is instituted, definitive evidence must be presented to administrators that the physical education program is indeed effective, and/or requiring increased instruction time.

Conversely, there are skeptics within the education system who question the contribution of physical education, according to Cooney (1987). Budget cuts within sports programs and the elimination of certain sports could be interpreted as evidence of a lack of appreciation for the benefits of physical education. It could be argued that other subject areas are receiving fewer resources also. However, considering that programs such as French are growing tremendously, to the point of daily instruction, it is apparent that these programs are viewed as somehow more valuable. Physical educators must disprove the skeptics by illustrating that physical education is, in fact, a beneficial discipline. This can best be accomplished by demonstrating how physical education programs are satisfying the mandate of the Ontario Ministry of Education.

Determining the effectiveness of the physical education instruction provided for Norfolk students is important for another reason. It could help answer the question, Are physical education teachers doing a good job? Answering this question affirmatively would have a positive effect on physical education teachers by confirming the value of their efforts.

The aforementioned issues can only be resolved after it has been determined whether the existing physical education program is effective. Evidence of an effective program would provide fuel for the movement toward daily physical education instruction; answer skeptics who would surrender physical education instruction time to other subject areas; and provide physical education teachers with positive feedback concerning their efforts. Should a program be determined to be ineffective, then a different course of action would be suggested; alternatives include curriculum and personnel review, and restating Ministry of Education objectives concerning physical education.

Methods of Evaluating Physical Education Programs

Physical education programs can be examined with respect to several different factors. Game skills acquisition and development, cognitive development, attitude development, or social skills development could have been used as the criterion of program effectiveness. However, a more tangible yardstick, physical fitness improvement, was chosen.

The Ministry of Education for Ontario (1978) stated that, "A valid and effective program in physical education is one that accomplishes its objectives. If an increased level of physical fitness is one objective, then evaluation of an effective program would reveal that the fitness level of a majority of the students actually did improve" (p. 20). Contained within this quote is a criterion by which the Ministry judges program effectiveness: fitness improvement in the majority of the population tested.

Pifer's (1987) definition of program effectiveness, while undefined, seemed to be very similar: she determined effectiveness when scores on ten of eleven fitness tests showed improvement. (There was no mention of the minimum standard for determining program effectiveness.)

When evaluating program effectiveness, some researchers considered success to mean simple improvement in fitness scores. In this research, significant improvement is necessary in order to conclude that the Norfolk physical education program is effective. This is not an unreasonable expectation when one considers the questions Tallmadge (1977) developed to provide evidence of program effectiveness. He asked two critical questions: 1) did a change occur?, and 2) was the effect statistically significant? By Tallmadge's standards, any physical education program which did not improve fitness levels significantly would be deemed ineffective with respect to that aspect of the program.

The need for standardized fitness testing

Over the last two decades there has been a growing realization on the part of both researchers and practitioners that the field of physical education requires a standardized fitness measuring tool. The Canadian Association of Sport

Sciences (1982), the Committee on Physical Fitness Research (1974), and the World Health Organization (1971) have each expressed this concern (Shephard, 1986), as have, more recently, the American Alliance of Health, Physical Education, Recreation, and Dance (1988), the Ontario Medical Association (1987), and the President's Council on Physical Fitness and Sport (1988). Unfortunately, this standardized measure has yet to come to fruition. Consequently, there remain a number of tests for every fitness component.

Anthropometric measurement of fitness

Anthropometric measurements of fitness could be used. These involve various measurements of the body intended to indicate the overall condition (fitness) of the participant.

Body mass is one such measure; the inference was that the greater the mass, the greater the degree of body fat, and consequently a lower level of fitness. However, high mass could also be an indication of a heavy frame or a well-developed musculature (Shephard, 1986), each of which is denser thus heavier than fat.

A height/mass index is in use also. It states the relationship between the height and mass of a subject, indicating the likelihood of obesity in the subject. This index could also be erroneous because of the body mass inaccuracy, that is, the fact that muscle tissue is heavier than fat tissue (Shephard, 1986).

Skinfold readings are used to determine percentage body fat.

This technique has one drawback, that being "...that determinations of sub-cutaneous fat are now realised to be heavily influenced by the technique of a given observer..." (Ross & Pate, 1987). Blair et al. (1983) and Shephard (1986) concur. With sixteen testers operating within this study the possibility of varying techniques, hence error, seemed high. For this reason, among others, skinfold readings were excluded from the test battery used in this research.

Lastly, body girths and body densities have been employed to determine fitness. The problem with this technique, as well as all of the other anthropometric measures listed, is that expensive equipment is needed and a considerable amount of time is required. In most cases, Norfolk schools simply do not possess the various scales and calipers required. In addition, one of the conditions for conducting this research was that time requirements of teachers would be minimal, and accurate caliper testing can be time-consuming. For the reasons enumerated above, fitness was determined through motor testing of health-related and skill-related components.

Defining Physical Fitness

It could be asked why fitness measurement should be done at all. CDDS (1983), Corroll (1983), Fox et al. (1987), and Hunsicker (1963) have all suggested that fitness testing must be a part of any successful physical education program. Beauchamp (1980), the Committee for the Development of Sport (1983), and

Elson (1981) have stated the multiple roles of fitness testing as: 1) determining the physical status of each student; 2) identifying children in need of special help; 3) measuring progress; 4) carrying out more meaningful planning; and 5) safety.

If physical fitness measurement is to be an integral part of any physical education program, and if physical fitness improvement is to be used in determining program effectiveness, it is necessary to establish an operational definition of physical fitness.

The definitions of physical fitness are as numerous and diverse as the researchers exploring the topic. Individual definitions tend to reflect the fields from which the researchers originate. Medical researchers might define a fit person as one who has "freedom from disability and disease" (Hunsicker, 1963, p. 5). Coaches, on the other hand, might consider fitness in terms of one's ability to perform demanding game-related tasks. These widely disparate views are indicative of the confusion generated when researchers attempt to pinpoint the term. Pate (1983) suggested that, "youth fitness has been operationally defined so many different ways that it has become meaningless" (p. 77).

Health-related versus skill-related fitness

Regardless of the many definitions circulating within the field of physical education, one trend is clear: there is a changing concept of fitness, "...paying special attention to distinguishing between physical fitness and athletic fitness" (Plowman, Hastad, & Marett, 1983, p. 46). This indicates that researchers are differentiating between fitness reflected in specific motor ability and that reflecting an overall sense of fitness. This change in thought is reflected in the definition of fitness used for this study: the ability to function in one's life without undue fatigue (Percival, Percival, & Taylor, 1977). The change in direction in fitness definition has resulted from the poor physical condition of the "average" American (Falls, 1980).

Several authors have acknowledged health-related fitness as critical (Beauchamp, 1980; Blair, Falls, & Pate, 1983; Falls, 1980; Pate, 1983). This is the type of fitness required of the average person, as opposed to skill-related fitness, which is essential for highly trained athletes. Falls (1980) stated that emphasis should be placed on developing health-related fitness, that is, fitness for everyone. His basis for this argument could be supported for a couple of reasons. First, development of programs for improving physical fitness for everyone is not only more democratic, but also more practical: what percentage of people achieve gifted-athlete status? Secondly, a person who is highly trained (athletically fit) in a specific activity might not necessarily be fit in a health context (Falls, 1980). This is not to say that performance-related and health-related fitness are completely unrelated. In effect, the difference is in degree (Beauchamp, 1980).

Delineation of Fitness Components

Cureton (1983) felt that one of the important gains in fitness test research has been the attempt to identify the components of fitness and to describe appropriate tests for measuring them. Unfortunately, the number of components delineated becomes unwieldy: balance, reaction time, flexibility, agility, strength, power, muscular endurance, cardiorespiratory endurance, coordination, speed, body composition, and more. With this large number of variables, a problem arises. When fitness components have been properly delineated, there is very little correlation between components, according to AAHPERD (1980), and Corbin, Dowell, Lindsey, & Tolson (1981), and Hunsicker (1963). For example, an elite distance runner might score tremendously high on a test of cardiorespiratory fitness and very low on a strength test. The converse is also possible: a superheavyweight weightlifter could score extremely high on a strength test but poorly on a test of cardiorespiratory fitness. This unrelatedness is to be expected if statistical analysis is considered. Factor analysis, a procedure often employed in fitness analysis, identifies the contribution of each fitness component to total fitness. Theoretically, individual contributions are unrelated, according to Bosco and Gustafson (1983).

Clearly, one of the reasons for identifying fitness components is in order to develop reliable methods for measuring each of them. This being the case, it behooves researchers to condense this list to a more workable size, for ease of test development and test administration. The problem is that short test batteries will not measure all fitness components (Hunsicker, 1963). The question then becomes, "which components should be included in a test battery of reasonable length?"

The answer to this question lies in the researcher's bias concerning the definition of physical fitness. There are two divisions of fitness: health-related and skill-related (Beauchamp, 1980; Corbin, Dowell, Lindsey, & Tolson, 1981; Falls, 1980; Pifer, 1987). Skill-related fitness is also referred to as performance-related (Corbin et al., 1981, p. 7) or motor-related (Pate, 1983, p. 77). Prior to 1979 most physical education researchers focussed upon the skill-related components of fitness. Based upon the research of Falls et al., the American Alliance of Health, Physical Education, and Research (AAHPER) developed a health-related fitness test (1980).

Components of health-related fitness

Health-related fitness, as the name implies, is "concerned with aspects of physiological functioning which offer some protection against degenerative type diseases such as heart disease, obesity, and musculoskeletal disorders" (Falls, 1980, p. 25). Caspersen (1987) and Smith and Gilligan (1987) concur.

The components of this division of fitness are generally agreed upon but there is no absolute consensus.

Corbin et al. (1981) and Falls (1980) considered strength, flexibility, body composition, and cardiovascular endurance to be the health-related components of fitness. Pifer (1987) substituted muscular endurance for body composition, while Fox (1987) included both of these alternatives. The components of health-related fitness have one thing in common: they contribute to the general wellness of everyone.

Strength is important in the protection of joints and the prevention of postural problems (Falls, 1980). This is especially true of abdominal strength, an item tested in the Canada Fitness Award battery.

Body composition deals with the relative amounts of fat and muscle in the body. The proportion of fat in the body is important because fat produces a drag effect on the body, forcing all systems to work harder to counteract its presence. The higher the fat content of the person, the more poorly the body performs (Greenberg & Pargman, 1986).

Flexibility, which is the ability to move the body through a range of motion, is important because it helps prevent exercise-related injury (Greenberg & Pargman, 1986).

Muscular endurance, the ability to do continuous muscular work, and cardiorespiratory endurance, the ability of the lungs to provide oxygen to the blood, are closely related because muscular endurance is dependent upon cardiorespiratory fitness (Greenberg & Pargman, 1986). They are both important in providing the individual with the ability to sustain activity in everyday life without undue fatigue.

Most researchers agree that cardiorespiratory endurance is the most important component of fitness (Falls, 1980), the reason being that cardiorespiratory fitness is critical in the prevention of circulatory disease, a major killer in our society (Greenberg & Pargman, 1986).

These health-related components provide the foundation for much of the fitness testing in the 1980's.

Components of skill-related fitness

Skill-related components of fitness include: coordination, speed, power, agility, and balance (Beauchamp, 1980). These attributes are not critical to the everyday well-being of the average person. Conversely, they contribute greatly to the ability to participate in athletic endeavours.

While physical education researchers have made the distinction between health-related and skills-related fitness, Beauchamp (1980) pointed out that there is a fine line between the two; the difference between the two is largely one of degree. Falls (1980) stated that the "essentials of fitness" (the health-related concepts) are also important components of performance-related fitness (p. 25), and Gabbard, LeBlanc, and Lowy (1987) suggested that the development of health-related components contributes to better performance in many skills-related activities. The overlap in definitions is also seen in the explanation of the value of flexibility (a health-related component) as prevention against exercise-related injury (Greenberg & Pargman, 1986). Much of the exercise could come from skill-related activity.

If the two divisions of fitness are so closely related, why are researchers determined to make a distinction between the two? Differentiating between the concepts of health-related and skills-related fitness, when the difference is slight, contributes further to the confusion in physical education terminology and fitness measurement.

AAHPERD has led the movement from performance-related to health-related fitness measurement. One of the benefits of this shift is due to the fact that health-related fitness components are more responsive to training (Pate, 1983) while some of the performance-related components, especially speed and power, are more matters of physical endowment (Murphy, 1986; Pate, 1983). Supporters of health-related fitness testing claim that genetics can have a confounding effect on any fitness testing (Blair, Falls, & Pate, 1983; Cowart, 1987).

The author tends to be skeptical about a total shift away from all performance-related testing. The President's Council on Physical Fitness, originated in the U.S. in the 1950's, had as its mandate the improvement of physical fitness in children. After thirty years of testing and results that have disappointed physical educators, AAHPER, the association responsible for the program development, changed the components tested to ones which are more susceptible to training. Results should theoretically be better but, in reality, will this mean that children are more fit? The issue is being clouded and in a few years, when the health-related shift has been completed, some faulty conclusions will have been forthcoming. Malina (1988) also questions the move totally away from skill-related fitness measurement, stating that there should be a balance between the two divisions of fitness. He feels that students are not being given the opportunity to develop motor skills.

In support of the emphasis upon health-related fitness as opposed to skill-related fitness, Blair et al. (1983) stated that within the AAHPER Youth Fitness Test (1976), forerunner to the Health-Related Test (1980), there was an inherent element of competitiveness. This, he felt, was wrong; individual fitness should be a personal matter. There should be no importance in comparing results. The Health-Related Fitness Test (1980) had its emphasis, according to Blair et al. (1983), on "achieving an optimum score that represents positive health status" (p. 94).

Within Blair's frame of reference, there must be a normative "score that represents positive health status". It must have been arrived at through use of test scores provided by a wide segment of the student population. This being the case, Blair was asserting that the emphasis for every student should be on the achievement of an optimal score attained by a given percentage of the population. This objective, competing with other students, does not differ significantly from what Blair claimed to be a fault in the Youth Fitness Test.

Arguments such as Blair's in no way invalidate the use of the Canada Fitness Award battery of tests. This battery was developed through the efforts of CAHPER, whose Fitness-

Performance Test is the most widely used test of fitness in Canada (Beauchamp, 1980) and has its roots in the AAHPER battery. The fitness components employed in that test battery are identical to the ones used in this study: cardiorespiratory endurance, muscular endurance, strength, agility, speed, and power.

Measurement of Fitness Components

Cardiorespiratory endurance

Assessing cardiorespiratory endurance is complicated because of the numerous components at work: the heart, the lungs, blood vessels, capillary system, and the oxygen capacity of the lungs (Bosco & Gustafson, 1983). Typical measurements could include: heart rate, stroke volume of the heart, blood pressure, and oxygen utilization.

Early attempts to measure this component "... were developed by medical doctors and consequently tended to accurately distinguish between healthy and unhealthy patients, but they lacked the sophistication to determine varying degrees of cardiorespiratory efficiency" (Bosco & Gustafson, 1983). These early tests were termed postural tests because they measured the effects of exercise or changes in body positions upon heart rate and/or blood pressure.

At the most modern, the testing of cardiorespiratory endurance is extremely technical, including such techniques and associated equipment as electrocardiography, heartography, ballistocardiography, and telemetry (Bosco & Gustafson, 1983). It is obvious that neither of these extremes is applicable to the modern elementary school scene.

During the 1940's, fitness tests, such as the Harvard Step Test, were evolving which required more cardiorespiratory efficiency than the earlier postural tests. Examples are the 12-Minute Run-Walk Test and the 15-Minute Run-Walk Test (Bosco & Gustafson, 1983). These tests have been highly correlated against maximum oxygen uptake, thereby establishing their reliability as tests of cardiorespiratory endurance (Bosco & Gustafson, 1983). The Canada Fitness Award test of this component, the endurance run, was chosen for its similarity to such tests in terms of elapsed time during the performance of the test.

Power

Power is often tested by the standing long (broad) jump. The Newton Motor Ability Test (1939), the Carpenter Motor Ability Test (1942), the Barrow Motor Ability Test (1954), the Scott Motor Ability Test (1959), and AAHPER (1976) have all employed this test of leg power. The CFA test battery likewise includes this test. This test is a measure of the distance covered by one double-foot hop. The inference is that longer jumps indicate greater power.

Speed

Speed can be measured by dashes, often in the 50 to 60 meter range (Barrow, 1954; AAHPER, 1976). The CFA battery includes the timed 50-meter dash as a measure of speed.

Strength

Strength and muscular endurance have been measured through use of push ups since the 1920's (Rogers Strength Inventory and Physical Fitness Inventory, 1926) and continues to be (AAHPER, 1976; Sharkey, 1979). Push ups, also a CFA test, are used in this research to measure muscular endurance.

Agility

Agility test items should include: change of direction in running, change of body position, and changes of direction of body parts (Bosco & Gustafson, 1983). The shuttle run satisfies these requirements and is used in the CFA test battery. Other tests to include the shuttle run are: the Phillips JCR Test (1947); the Fleishman Physical Fitness Test Battery (1964); AAHPER (1976).

Muscular endurance

Curl-ups satisfactorily measure abdominal strength and endurance (AAHPER, 1976; Sharkey, 1979) and were used as part of the testing for this research. They are also a component of the CFA battery. Sit-ups have been a more traditional test, however that test does not measure abdominal strength solely, and are considered to be inferior to curl-ups (Jette et al., 1984). (Whereas sit-ups generally require the participant's hands to be locked behind the head, curl-ups require the hands to remain on the thighs throughout the exercise.)

Factors affecting fitness testing

There are factors which can influence any fitness testing. According to Meyers (1980), 1) insufficient test orientation, 2) inadequate test preparation, 3) inaccurate scoring, 4) inadequate rest between tests, and 5) lack of willingness to put forth best effort, can all act to render test results less valid than possible.

The first four factors can be controlled with reasonable care. If the tester is made familiar with the test battery well in advance of the proposed testing period; if responsible helpers (students or adult volunteers) are selected; and if common sense is used, the first four worries can be eliminated. However, "lack of willingness to put forth best effort" is a definite concern that has been identified by other researchers, such as Shephard (1986). While the Committee for the Development of Sport (1983) felt that fitness testing itself acts as a motivational aid, this view does not represent the consensus of thought.

The effects of motivation are obvious, especially in the area of physical performance. It could be suggested that if, on a fitness pre-test, a student does not try his/her best, the results will be poorer than those on a post-test, where the student was in a mood to try hard. The treatment (physical education instruction) may not have had any effect at all, but the appearance is just the opposite. It could be argued that there exists in every person a potential for physical performance, the limits of which no one knows, not even the individual concerned. If physical education does nothing but motivate the student to more closely approach that physical potential, then the physical education program is a benefit. Raithel (1987) addressed the need for girls, especially, to more closely approach their physiological potential.

A variable closely related to motivation is encouragement. Whereas encouragement is the external influence of those around a person, motivation is an internalized desire to do well, possibly, but not necessarily, the result of encouragement. Shephard (1986) stated that fitness test scores are heavily dependent upon encouragement. This being the case, if one tester uses excessive encouragement while another uses none at all, and if the students involved are equally skilled, the results would be disparate in favour of the student receiving the encouragement. This is not to suggest that the effects of encouragement are powerful enough to overcome superior natural fitness, however it is reasonable to assume that this variable could have a confounding effect on fitness testing. Fitness Canada (1985) acknowledged this potential problem by suggesting that no encouragement be given at all during testing. Following that instruction eliminates the confounding effect of encouragement within this research.

Criterion-referencing versus norm-referencing

Often, when measuring fitness, individual scores are compared to standardized scores. These standards are developed as a result of large numbers of subjects; in effect, standardized scores are expectations of performance. The expected scores, or norms, are misleading, according to Blair et al. (1983), because they are often expressed as percentiles. For example, an individual might score at the 20th percentile, which means that 80% of a population scored higher than that individual. This would suggest an unfavourable performance. However, if the population were abnormally gifted with respect to this particular characteristic, the individual might actually possess an adequate amount of the characteristic. Conversely, an individual might score at the 80th percentile while not really possessing an abundance of the characteristic being tested. The entire population, 80% of which is poorer than the individual, might be generally lacking in that particular characteristic. Also, when computing fitness improvement, percentiles are misleading because a five-percentile improvement at the 85th percentile is more significant than a five-percentile improvement at the 45th percentile.

Blair (1983) suggested that criterion-referenced standards are preferable to norm-referenced standards. Criterion referencing suggests that an individual should be judged according to the amount of the characteristic possessed, not by a comparison with others. One concern with this form of referencing is that an expert is needed to determine what level of a particular characteristic is adequate. Considering that experts might disagree, the difficulty is obvious: who establishes the criterion?

The Canada Fitness Award employs percentiles. These reflect the scores of the nineteen million students who have participated in the test (Fitness Canada, 1989). With this large number of participants, a normal distribution can reasonably be assumed. Also, the performance levels for awards are chosen by experts. In effect then, the CFA uses both norm referencing (percentile scores) and criterion referencing (choosing appropriate levels for awards). For this reason, the CFA standards are useful for comparison.

Fitness improvement was not calculated in terms of percentile improvement, for the reason given by Blair (1983). Instead, each student was compared with him/herself.

Fitness Test Batteries

Every researcher must determine which fitness components are important, then reliable tests of those components must be defined. Once accomplished, a test battery has been defined, the purpose of which is to describe the overall fitness of an individual. The combining of individual fitness scores is desirable and convenient in a descriptive sense; the same was attempted in this research. However, Corroll (1983) offered a caution when attempting composite scores, especially averages. When composite scores are determined, weaknesses in particular test areas are often overlooked. Considering that "identification of children in need of special help" (p. 7)is an objective of fitness testing (Beauchamp, 1980), Corroll's concern seems warranted. An alternative, according to Corroll, would be individual profiles for each test item. That is also attempted in this study.

There are several fitness testing batteries in existence: the Carpenter Motor Ability Test (1942); the Phillips JCR Test (1947); the Barrow Motor Ability Test (1954); the Scott Motor Ability Test (1959); the Fleishman Test Battery (1964); the Canada Home Fitness Test (1967); the AAHPER Youth Fitness Test (1976); the AAHPER Health-Related Fitness Test (1980); Eurofit (1983); the Canada Fitness Survey (1984); the Canada Fitness Award (1984). A new AAHPERD-sponsored fitness battery, "Physical Best" is scheduled for implementation in the 1988-89 school year (Raithel, 1988). Each test battery carries an inherent bias, based largely upon the date of inception. Earlier batteries focussed upon performance-related fitness concepts while recent ones are more concerned with health-related concepts. The confusion caused by the multitude of test batteries, and concomitant biases, has renewed the call for a universal battery of fitness tests. In the U.S., there has been a call for a single test that AAHPERD and the President's Council on Physical Fitness could both promote (Murphy, 1986). "Physical Best" is intended to address that need, although acceptance of it is confusing at this point (Raithel, 1988).

Eurofit, a European version of AAHPER and CAHPER, is the test battery developed under the auspices of the Committee for the Development of Sport (1983). This battery is extremely close in nature to the CAHPER-developed tests. The fitness components identified by Eurofit are: cardiorespiratory endurance, strength, muscular endurance, power, flexibility, speed, and balance. In addition, the tests are designed for ages six to eighteen, very close to the six-to-seventeen range of the Canada Fitness Award battery.

The importance of Eurofit is that it represents an international attempt to standardize fitness testing. Cuniff (1985) might have been overly optimistic when he claimed that by 1990 there should be "a methodology for systematically assessing the physical fitness of children" (p. 16), but the intent is clear.

The Fitness Battery Used in This Study

In this research the Canada Fitness Award battery of tests is the measurement tool. The components measured by this battery are: muscular strength and endurance (abdominals, arms, and shoulders), agility, speed, cardiorespiratory endurance, and power (Fitness Canada, 1985). The choices of muscular strength, muscular endurance, and cardiorespiratory endurance are easily justifiable since virtually all fitness batteries include them (Bosco & Gustafson, 1983). Agility and speed are also included in several researchers' choices of definable fitness components (Hunsicker, 1963; O.A.C. Validation Draft, 1988; Ontario Ministry of Education, 1978). Other researchers have chosen the identical fitness components to the CFA for consideration (CDDS, 1983; Percival, Percival, & Taylor, 1977; West, Beveridge, & Workman, 1987). The CFA fitness components chosen also satisfy the dilemma of health-related versus performance-related fitness. The three universally accepted components (strength, muscular endurance, and cardiorespiratory endurance) are health-related, while agility, power, and speed are performance-related.

Falls (1980), Beauchamp (1980), Corbin et al. (1981), and Cuniff (1985) felt the emphasis should be on health-related fitness, that is, fitness for everyone. In reality, the emphasis in schools has been on performance-related fitness (Corbin et al., 1981; Cuniff, 1985). An advantage of the CFA battery of tests then is that it includes both performance-related and health-related fitness measures.

The two health-related fitness components which are not included in the CFA battery are flexibility and body composition. These two components offer the elementary school teachers distinct problems. Both require measurement devices, in the case of flexibility a Sit-and-Reach type device and, in the case of body composition, calipers. Very few, if any, elementary schools in Norfolk have access to such devices. An additional drawback of caliper use is the problem of inaccuracy (Ross & Pate, 1987).

The Canada Fitness Award is a "fitness incentive program for Canadians from six to seventeen years of age" (the Inside Story, 1985, p. 3). It is concerned exclusively with motor fitness and draws upon the research of the Canadian Association for Health, Physical Education, and Recreation (CAHPER).

The test battery consists of: 1) push ups, 2) shuttle run, 3) partial curl-ups, 4) standing long jump, 5) 50-metre run, and 6) an endurance run.

These tests satisfy the CAHPER guidelines for selecting physical performance tests: i) test validity and reliability (Lashuk, 1984), ii) easy administration, iii) minimal equipment and time, and iv) common items with other fitness testing programs to facilitate comparisons (Fitness Canada, 1985).

Awards are presented on the basis of performance on all test items. An Excellence rating on an individual test is presented to any student who achieves the 85th percentile; a Gold rating is awarded to those attaining the 75th percentile; Silver is awarded to those reaching the 50th percentile; and Bronze the 15th percentile.

An Excellence crest is awarded to those who achieve the Excellence level on all six test items. A Gold crest is awarded to students achieving the Gold level on five test items, including the endurance run. A Silver crest is presented to those attaining the Silver level on four test items, including the endurance run. A Bronze crest is awarded to those students achieving the Bronze level on four test items, including the endurance run (Fitness Canada, 1985). (The critical condition for earning any crest is performance on the endurance run. It could be suggested that this fact indicates Fitness Canada's appreciation of the critical importance of cardiorespiratory endurance, however Ian Craigon, director of CAHPER'S Quality, Daily Physical Education (1989), has stated that the aforementioned condition for earning crests reflects the desire of some Canadian physical educators to stress cardiorespiratory conditioning in schools. Crests earned were not used as a criterion of program effectiveness because of this weighting of components).

While the CFA battery of tests was employed mainly as a data-gathering tool for this study, the various achievement levels are important for: comparison of Norfolk and national averages in both mean performance scores and numbers of awards, and assigning values in order to calculate aggregate fitness scores.

Determinants of Fitness Performance/Improvement

Physical activity

Based upon past research, the benefit of physical education lies in the physical activity which is offered to students. Numerous researchers (Gabbard, LeBlanc, & Lowy, 1987; Greenberg & Pargman, 1986; Hashel, Montoye, & Orenstein, 1985; Malina, 1988; Ontario Medical Association, 1987; Ross & Pate, 1987) have cited a strong positive relationship between fitness improvement and participation in physical activity. One dissenter to this view is Shephard (1986), who claimed that from his research this relationship was not always strong.

A possible explanation for the difference of opinion between Shephard and the supporters of the value of physical activity is that researchers often determine the activity levels of students based upon participation in organized, categorized activities. Younger children, especially, spend hours running, skipping, climbing, and wrestling, but when surveyed, their parents might answer that their children are not involved in the physical activities listed. The conclusion might erroneously be reached that these are inactive children. Several researchers have acknowledged this conundrum, including Pate and Ross (1987).

Activity levels among children have substantially decreased over the years (Cooney, 1987). Pate and Ross (1987) and Raithel (1987) have claimed that the amount of time spent watching television is directly related to lower activity levels in children. In effect then, these researchers are holding television, and other passive forms of entertainment, partially responsible for the decreased levels of fitness exhibited on recent fitness tests. This variable should be included in a more comprehensive examination of fitness levels.

The diminution of time available for physical activity is critical when the side effects are considered. With decreased activity come higher levels of body fat (Blair, Falls, & Pate, 1983), which, in children, can result in the pre-conditions for higher instances of circulo-respiratory ailments as adults. Ross and Pate (1987) identified participation in regular exercise programs as one of the best ways of lowering the risks of degenerative diseases. Other researchers concur (Caspersen, 1987; Rippe, 1987; Smith & Gilligan, 1987).

Parental attitudes

Like so many other value systems, children's attitudes toward physical activity are apparently influenced by their parents. McGinnis (1987) stated that the health-related fitness of children is significantly related not only to the activity levels of the children, but also to the activity levels of the parents. Children learn young what their parents think about physical activity and, in most cases, model that behaviour.

Gender

Gender can be a determinant of fitness performance. Over the last fifteen years, fitness performance has improved most dramatically among Canadian adolescent girls (Gauthier et al., 1983). The most logical explanation of this phenomenon is that equality is reaching amateur athletics; no longer is society suggesting that females are not fit for athletics, and vice versa. However, the cycle of change has not been completed: twice as many boys still participate on sports teams (Ross & Pate, 1987). Increased participation and performance notwithstanding, Canadian females continue to score poorly on tests of cardiorespiratory endurance, strength, and muscular endurance (Alexander, Ready, & Fougere-Mailey, 1985).

It seems ironic that females would score most poorly on tests of health-related fitness, rather than performance-related fitness, considering that what females have lacked most is exposure to sports activities. Shephard (1986) adds to the quandary by stating that girls often achieve a better muscular endurance than do boys, peaking at thirteen or fourteen years of age.

Female fitness-performance scores also cause consternation for a physiological reason: girls show a less sharp growth rate than do boys (Shephard, 1986). This growth rate peaks between eleven and thirteen. It could be suggested that it should be easier for females to become accustomed to functioning with their more moderately growing bodies. They should less frequently exhibit the gangliness of adolescent boys, and Shephard (1986) has stated that performance tests are affected by body size. Corbin (1980) offered partial explanation for the questionable performance of females when he suggested that adolescent girls could probably achieve more if they really wanted to.

Resource investment

Pate and Ross (1987) have suggested that the resource investment of the school has an impact on fitness components, most notably cardiorespiratory fitness. School resources include, among other things: 1) instruction time offered for physical education, and 2) qualified physical education teachers. The value of increased physical education instruction time has been established already, and Pate and Ross (1987) conclude that the physical fitness of students is related to the qualifications of the physical education teacher.

Natural physical endowment

Several researchers, including Shephard (1986) and Cowart (1987), have noted that natural endowment of physical skills is a determinant of fitness performance. This is always a confounding variable for teachers; some students, even without instruction, are naturals at some educational activity, be it math, spelling, or running. Bearing this in mind, it is important, when attempting to measure the effectiveness of a physical education program, to initiate fitness tests early in the year in order to establish base fitness levels. The September pre-test used in this research functions as that fitness baseline.

Initial fitness level

This initial fitness-performance level is a consideration for another reason. Shephard (1986) and Pate and Ross (1987), among others, have concluded that more exercise is needed in order to improve the fitness level of an individual whose initial condition is good, than for an individual whose fitness level is low. This means that it should be easier for poorer athletes to improve their physical conditions, assuming they had the desire to do so.

Motivation

The fact that it should be easier for poorer athletes to improve fitness levels does not necessarily translate into proportionately more greatly improved performances however. Several researchers (Beauchamp, 1980; Gabbard, LeBlanc, & Lowy, 1987; West, Beveridge, & Workman, 1987) have acknowledged that motivation is an ever-present factor to be considered in performance testing. Greenberg and Pargman (1986) consider motivation even more important in the performance of children that of adults.

In summary, then, research has shown that gender, motivation, activity levels, school investment in physical education, and natural endowment are determinants of fitness performance and, arguably, fitness level improvement.

The Contextual Framework for this Study

In order to place this research in context, it is necessary to briefly relate this study to similar work. There are few studies which specifically set out to examine the effectiveness of physical education programs. However, while examining the change in fitness performance, the objective of many of these studies, the drawing of conclusions is inevitable. The conclusions drawn concerning program effectiveness rely upon the definition of that concept used by the researcher. This research differs from other studies in the accepted definition of program effectiveness; the definition used in this research is more stringent than those used by the Ministry of Education (1978) or Pifer (1987), although not unsupported (Tallmadge, 1977).

The measurement instrument in this study again varies from present convention. While recent American research is tending toward health-related fitness measurement (Falls, Pate, & Ross, 1983), this study is similar to European research (CDDS, 1983) which provides a measurement tool concerning itself with performance-related fitness, as well as health-related fitness.

The tests of each fitness component are quite traditional, with the exception of the curl-up. This exercise has replaced the sit-up as a test of abdominal muscular endurance because it is more muscle-specific than the sit-up, which, besides measuring abdominal endurance, also measures endurance of the hip flexors.

Summary of Related Research

In summary, the current literature suggests that one of the purposes of evaluating any educational program is to determine the effectiveness of that program. Several methods exist for evaluating physical education programs, but fitness measurement is the most common. In order to measure fitness, an operational definition of physical fitness must be chosen. Physical fitness is a multi-dimensional concept, consisting of several independent components which must be identified. Appropriate tests exist for the measurement of each component.

Performance on these fitness tests is related to certain factors, most notably gender.

CHAPTER III

METHODOLOGY

The Research Setting

Situated along the north shore of Lake Erie, between Haldimand and Oxford Counties, Norfolk County is a predominantly rural area of 1600 square kilometers; Norfolk and Haldimand Counties amalgamated in the early 1970's to create the Regional Municipality of Haldimand-Norfolk, although there are separate education systems for each county. Simcoe is Norfolk's largest community, with a population of 15000, while the remaining larger centres are Port Dover (4000), Delhi (5000), and Waterford (2500).

The Norfolk Board of Education, with 15 board members, serves this area with five secondary schools (Port Dover Composite School, Simcoe Composite School, Delhi District Secondary School, Waterford District High School, and Valley Heights Secondary School) and 21 elementary schools. Thirteen of these elementary schools have grades seven and eight students, and twelve of these schools agreed to participate in this research: Bloomsburg, Boston, Courtland, Delhi, Doverwood, Elgin, Houghton, Langton, Nixon, Teeterville, Walsh, and Walsingham.

The Director of the Norfolk Board is assisted by four superintendents, one of whom is Superintendent of Programs, School Organization, and Planning. Pertinent to this research, there is also an Intermediate Supervisor of Educational Programs, although there is no consultant for physical education specifically.

There are 16 teachers of physical education in grades seven and eight. Some have physical education specialization, while others have none at all. That discrepancy in training plays a major role in the findings of this research.

Chapter Overview

This chapter describes the research hypotheses for the research, as well as the design and methodology used to determine the validity of those hypotheses. Included are the data collection tool, field procedures, methodological assumptions, and limitations.

Students performed the Canada Fitness Award (CFA) battery of tests in September and again in May. The September results established the initial fitness levels of the Norfolk students. The May scores established the final fitness levels of the students for the test period. The two sets of data were initially used independently, the only concern being "are Norfolk students fit presently?".

A comparison of results later formed the basis for determining program effectiveness.

The research design is descriptive as no control group was used. The evaluative nature of the research, as opposed to pure experiment, allows for the use of this design.

An experimental design could potentially have proved a more powerful tool, however, there was no appropriate control group available. An experimental, daily physical education program might have been established. However, implementation of such a program would have exceeded the extent of Norfolk Board of Education involvement in this research. Also, since Norfolk scores were compared with CFA norms, it could be assumed that these norms could be used as the control group in an experimental design, but standardized scores should not be used as a control group in an experimental design (Campbell & Stanley, 1963). The reason is that there can be no guarantee that the characteristics of the control group and experimental group are identical.

Another possibility could have been to randomly create both experimental (no physical education instruction) and control (physical education instruction) groups from within the population of Norfolk elementary students. The drawback to this approach was that physical education is considered a mandatory subject, so withdrawal of instruction in this subject area would have been inappropriate.

Population and Sample Variables

In terms of research design the dependent variable for this research was fitness-level change with gender, age, grade, teaching experience, instruction time, extracurricular physical activity, nature of the class, physical education specialization, and initial fitness level being independent variables.

Gender, age, grade, and teaching experience are self-explanatory variables.

Physical education specialization involved the possession of a degree in physical education, a physical education specialist's certificate, or some other specific physical education training.

Initial fitness level related to the aggregate fitness level of the student in September, as determined by quartile rankings.

Instruction time referred to the total minutes of physical education instruction per six-day cycle.

The nature of the class dealt with the matter of segregated versus coeducational classes.

Extracurricular physical activity involved any regular physical activity outside of physical education classes.

The null hypotheses, stated alternately from the research hypotheses listed in Chapter 1, were:

H_0#1 There will be no significant difference between the average Norfolk fitness scores and the average Canada Fitness Award scores. (The Norfolk fitness scores were calculated using mean scores while the CFA scores represented the 50th percentile. In a normal distribution the mean and 50th percentile represent identical points, thereby allowing for comparison.)

H_0#2 There will be no significant difference between pre- and post-treatment mean scores for the students on any of the tests.

H_0#3 There will be no significant relationship between fitness-level improvement and any of the independent variables tested.

With respect to H_0#1, there was a mean fitness score calculated for each subtest according to age and gender. Hence, there was: a mean curl-up score, a mean push up score, a mean 50-metre run score, a mean shuttle run score, and a mean endurance run score for each of eleven-year-old boys and girls, twelve-year-old boys and girls, thirteen-year-old boys and girls, fourteen-year-old boys and girls, and fifteen-year-old boys and girls. Each of these scores was compared to CFA standards.

In addition, the raw scores for each test were converted to point values, ranging from zero to four, corresponding to the CFA level achieved. When combined, an aggregate fitness score resulted. These aggregate scores were also compared with CFA norms; using the silver level (50th percentile) multiplied by six (the number of subtests) resulted in a score of twelve being the CFA standard for fitness.

Finally, the September scores were examined for relationship with another variable: summer physical activity.

In order to address H_0#2, the September scores were compared to the May scores. The identical items were compared as with H_0#1. In this case, however, the concern was not entirely in the standing of Norfolk students as compared to national standards, but also with comparing the same students over time.

H_0#3 was addressed by examining performance variance for relationship with age, gender, grade, teaching experience, instruction time, extracurricular physical activity, nature of the class, physical education specialization, and initial fitness level.

Pilot Studies

The Canada Fitness Award battery of tests had been sufficiently field tested to be considered a reliable tool for data collection (Lashuk, 1984). In addition, with over nineteen million subjects as of 1988 (Fitness Canada, 1989), the distribution of scores can arguably be assumed to be normal. This is a necessity when considering use of parametric statistics and when assuming that the 50th percentile point is identical to the mean of a distribution of scores.

Selection of Subjects

The population of Norfolk grade sevens and eights is approximately 1000. With one school refusing to participate and approximately eighty score sheets returned incomplete, the total number of participants was 885 in September.

Instrumentation

Raw performance scores were generated using the six Canada Fitness Award tests. These scores were used in answering the descriptive question featured in this study, "Are Norfolk students fit?"

Scores on each test were compared with the age- and gender-appropriate Canada Fitness Award standards. Points were awarded in accordance with the Bronze, Silver, Gold, and Excellent ratings: one point for the Bronze level; two points for the Silver level; three points for Gold; and four points for Excellent. No points were awarded when the minimum level was not attained. With six tests the maximum score possible was twenty-four and zero the minimum.

The Silver level of the CFA corresponds to the 50th percentile; hence, an average fitness level. Two points being awarded for attaining the Silver level, an aggregate score of twelve was considered to represent an average fitness level, and therefore a fit student (by CFA standards).

An alternate standard by which a student's fitness could be gauged would be to use the Silver crest level awarded by the CFA. The Silver crest is presented to any student who achieved the Silver level or better on four of the six CFA tests, including the endurance run. The author considers this method of crest achievement as inequitable. The Silver crest could be awarded to a student earning only eight total points on the six tests, while a student earning twenty points could possibly get no crest at all. This discrepancy is due to the condition attached by the CFA concerning the cardiorespiratory component of fitness.

It could be suggested that any awarding of points should be weighted to reflect the importance of the endurance run but more confusion would be created by attempting an appropriate weighting factor. The author found no examples of fitness test scores weighted in favour of cardiorespiratory endurance, or any other component of fitness. In addition, Fitness Canada did not suggest any weighting of components (Craigon, 1989). In fact, the inclusion of the condition of level-appropriate performance is the result of a request by physical educators, not an acknowledgement of Fitness Canada's belief that the cardiorespiratory component of the testing should be given added weight.

After determining the number of students who were considered fit (an aggregate score of twelve or above), this number was compared to the expected number as determined by the CFA. (Since the Silver level corresponds to the 50th percentile, according to CFA standards, 50% of all participants should score below the Silver level of fitness, 50% above. Therefore, 50% of all students should have scored twelve or above.) A Chi-Square test was used to decide whether the actual number of fit students was significantly different than the national averages.

In May, the identical manipulations were performed with the raw test scores. In addition, the mean scores for September and May were compared in order to determine if there was a significant difference between the two.

Field Procedures

Internal validity was a concern in this research, as in any other. In order to ensure that the data were valid, several precautions were taken.

Testers were advised to create testing conditions in September and May which were as similar as possible. These conditions included: time of day for the testing, field conditions, weather, and time framework. Testing to be done outside was to be completed in the afternoon in order eliminate wet and slippery conditions. Extreme heat or cold, and rainy conditions were to be avoided to provide optimal opportunity for top performance. Testing was carried out within two-week periods in both September and May.

The CFA testing package was previewed with testers during a group meeting in September. Testers were reacquainted with the test items, especially those added in 1984 (curl-ups and push ups). As a follow-up, the CFA test instructions were sent to each tester.

Data Collection

To ensure anonymity throughout the study, each participating student and school was assigned an identity number to be used for both sets of tests. The student numbers ranged from 0001 to 0885. The schools were numbered from 01 to 12. The student numbers were grouped according to school. For example, School #1 had student numbers 0001 to 0022; School #2, numbers 0023 to 0082, and so on.

Each student had a numbered information/performance sheet which contained the raw scores for the six CFA tests, the points awarded for those scores, the aggregate fitness score, and the quartile ranking for the student. In addition, pertinent demographic information was included, such as age, gender, grade, whether or not the student participated in a summer physical activity, and whether or not the student participated in an extracurricular physical activity during the interval between pre- and post-tests.

Teachers were also surveyed concerning certain characteristics. Factors considered were: teaching experience; possession of a physical education degree, a specialist's certificate in physical education, or some physical education specialization; the nature of the class (segregated or co-educational); and instruction time per cycle.

Analysis

Chi-Square Tests were conducted in September and May in order to determine if Norfolk students were fit; expected numbers of fit students were compared to observed numbers. Two-by-two tables were used to express the expected and actual numbers, thereby establishing 3.84 as the critical value for determining significant differences between observed and expected values.

Methodological Assumptions

There are methodological assumptions inherent in this research.

1. Testing conditions were similar enough to be an insignificant factor in explaining performance variance. Since these conditions were dependent upon the professionalism of the testers, this assumption seems valid.

2. It was assumed that the Canada Fitness Award satisfactorily identified the components of fitness. There seemed little doubt about the appropriateness of including measures of cardiorespiratory endurance, muscular endurance, and strength; their inclusion in fitness tests is virtually universal. The

inclusion of measures of agility, speed, and power was reasonable according to the fitness components chosen by the CDDS (1983), and included in its comprehensive fitness-testing package, EUROFIT. The Ministry of Education for Ontario (1978; 1986) also included these measures.

3. It was assumed that the CFA subtests were reliable measures of the identified components of fitness. This fact was supported in the research of Lashuk (1984), who stated that indeed the Canada Fitness Award was a reliable and valid fitness measurement tool.

Limitations

There are limitations in any research where assumptions must be made. In this research the limitation dealt with motivation. As many researchers (Bozzo, 1983; Gabbard, LeBlanc, & Lowy, 1987; Greenberg & Pargman, 1986) have observed, this factor requires consideration. Corbin (1980) stated that females in general appear to be less motivated in the area of physical activity and fitness. Harrer (1983) also noted that motivation was a concern with both students and teachers alike. It is apparent that, while several teachers expressed genuine interest in this research, it meant far more to the author than to any of the other teachers involved.

The testing might have had very different effects upon the students. Some might have been enthusiastic while others performed perfunctorily. (The CDDS felt that physical fitness testing acts as a motivational factor in and or itself, but most researchers do not conclude the same.) The issue in question here was, "Do the results of the tests indicate the true fitness levels of the students?". For the purpose of this research the answer must be in the affirmative.

A second limitation was the time available for testing of the students. As instructed by the Director of the Norfolk Board of Education, a minimum of class time was used for this research. Had this not been the case, a more comprehensive analysis of the results could have been forthcoming, involving information about parental values, specific extracurricular physical activities, and percentages of class time given over to various activities.

CHAPTER IV

FINDINGS (ANALYSIS AND EVALUATION)

The descriptive question featured in this research was, "Are Norfolk students fit?". In order to address this question, Norfolk fitness scores were compared with Canada Fitness Award (CFA) standards. Using Chi-Square Tests, it was possible to establish whether performance differences were significant.

With five CFA standings (Excellence, Gold, Silver, Bronze, and "below Bronze"), a Chi-Square test could determine the significance of performance variance, however, there could be no conclusion about whether Norfolk scores were generally better or worse than expected, and that was an objective of this research. To correct this situation, the five CFA standings were condensed into two: "above standard" and "below standard". Using the point system devised for this research, and described in earlier chapters (Excellence-4, Gold-3, Silver-2, Bronze-1, and below Bronze-0), "above standard" denoted an aggregate score of 12 to 24, while "below standard" included all scores from 0 through 11. In this manner, it was immediately evident where any variance existed, and it was possible to state that a group's performance was better or worse than expected.

The results were divided into three sections. First, there was a presentation of the September pre-test results. In general, these results established a baseline for later analysis. However, the results were also laid out in a fashion that allowed for analysis of specific independent variables. For example, the September data were presented with respect to gender, age, grade, school, and extracurricular physical activity. The second section was a repeat of all September testing in order to establish the May data. The third section was an analysis of performance variance, using T-tests and Chi-Square Tests to examine fitness change with respect to the independent variables already listed, in addition to teacher qualifications, teaching experience, and instruction time.

Fitness Performance and Gender

As can be seen on Table 1, there was no relationship between total fitness and gender in September. The Chi-Square value of 0.06 fell short of the critical value of 3.84 (df=1). This means that there was no significant difference in the performance of Norfolk males and females on the CFA tests, given the established standards for each gender.

Again in May, there was no relationship between gender and fitness performance (Table 2).

After comparing Norfolk females and males, their scores were compared with national standards. Table 3 illustrates the total fitness of Norfolk students, according to CFA standards. In short, Norfolk students were significantly unfit (.01 level) in September.

When categorized according to gender, Norfolk males performed only slightly better than females. While both genders proved to be significantly unfit, the females were even more so.

Table 1

Fitness and Gender (September)

	Males		Females	
	f_o	f_e	f_o	f_e
Above Standard	184	182.05	183	184.95
Below Standard	255	256.95	263	261.05

Chi-Square = 0.06.

Table 2

Fitness and Gender (May)

	Males		Females	
	f_o	f_e	f_o	f_e
Above Standard	228	225.51	207	209.49
Below Standard	180	182.49	172	169.51

Chi-Square = 0.13.

Table 3

Comparison of Actual and Expected Aggregate Fitness Scores (September)

		Above Standard		Below Standard	
Age	Gender	f_o	f_e	f_o	f_e
11	F	22	30.5	39	30.5
11	M	25	27	29	27
12	F	97	104	111	104
12	M	79	94	109	94
13	F	58	79.5	101	79.5
13	M	68	81	94	81
14,15	F	6	9	12	9
14,15	M	12	17.5	23	17.5
All	F	183	223	263	223
All	M	184	219.5	255	219.5

Chi-Square (Females) = (14.35), df = 1, $p < .001$.

Chi-Square (Males) = (11.48), df = 1, $p < .001$.

Note. Brackets in tables indicate scores which are poorer than expected.

When each gender was subdivided by age (Table 3), it was evident that every age group for both genders was generally unfit. Three groups (11-year-old girls, 12-year-old boys, and 13-year-old boys) were significantly unfit at the .05 level, while 13-year-old girls were significantly unfit at the .001 level. From this table, it can be stated that 11-year-old boys were the fittest group in Norfolk, while 13-year-old girls were the least fit. This statement does not suggest that the raw scores of 11-year-old boys were superior to all other groups; aggregate scores reflect comparisons with age-appropriate national standards.

Table 4 illustrates that, whereas in September every age and gender group performance was substandard, in May each of the aforementioned groups scored above standard. Twelve-year-old girls and thirteen-year-old boys scored significantly better than expected (.01 and .05, respectively). Males in total scored significantly better than expected (.05 level). In short, then, the students of Norfolk were fit in May (.005 level).

There was a noticeable change in the physical fitness levels of Norfolk students over the course of the year. Of the 785 students tested in May, 595 improved their fitness performances over September (Table 5). This represented a significant improvement in fitness levels (.001 level).

As can be seen on Table 5, significantly more males improved their fitness levels than did females, although improvement for both genders was significant (.001 level).

An aggregate score is the sum of the scores of each of the

Table 4

Comparison of Actual and Expected Aggregate Fitness Scores (May)

		Above Standard		Below Standard	
Age	Gender	f_o	f_e	f_o	f_e
12	F	74	59	44	59
12	M	55	51	47	51
13	F	89	89	89	89
13	M	107	93	79	93
14,15	F	44	41.5	39	41.5
14,15	M	66	60	54	60
All	F	207	189.5	172	189.5
All	M	228	204	180	204

Chi-Square (Females) = 3.23.

Chi-Square (Males) = 5.64, df = 1, $p < .05$.

Table 5

Change in Aggregate Fitness Scores from September to May

Overall

	f_o	f_e
Improved	595	392.5
Not Improved	190	392.5

Chi-Square = 208.94, df = 1, p < .001.

Females

Improved	261	189
Not Improved	117	189

Chi-Square = 54.86, df = 1, p < .001.

Males

Improved	334	203.5
Not Improved	73	203.5

Chi-Square = 167.38, df =1, p < .001.

six CFA tests. When the individual tests are considered, the inherent strengths and weaknesses of the genders can be identified (Table 6). On three of the tests there were significant differences by gender in the performances: distance run (.05 level), curl-ups (.05 level), and standing long jump (.01 level). On the distance run, the significant performance variance was due to the fact that boys scored more poorly than expected and girls scored higher than expected. On curl-ups, the significance in performance variance was due to the better than expected performance by boys and the poorer than expected performance of girls. As with the distance run, the poorer performance by boys on the standing long jump accounted for the significant performance variance.

For three of the tests there was no relationship between gender and performance (50-meter dash, push ups, and shuttle run).

Table 7 illustrates that there was also little gender difference in performance when four of the CFA subtests were considered separately in May. On the two remaining tests, standing long jump and curl-ups, however, there was a significant relationship (.001) between gender and performance.

After comparing the male and female results on the six tests among themselves, they were compared with CFA standards to determine how Norfolk students fared nationally on each of these tests.

In the standing long jump, as seen in Table 8, Norfolk

Table 6

Gender and Performance on Individual Canada Fitness Award Subtests (September)

Subtest	Chi-Square
Standing Long Jump	9.45***
50-Metre Dash	0.72
Curl-ups	5.43*
Push Ups	1.13
Distance Run	4.88*
Shuttle Run	0.79

*$p < .05$. ***$p < .005$.

Table 7

Gender and Performance on Individual Canada Fitness Award Subtests (May)

Subtest	Chi-Square
Standing Long Jump	20.84****
50-Metre Dash	1.51
Curl-ups	15.76****
Push Ups	0.06
Distance Run	0.46
Shuttle Run	0.00

****$p < .001$.

Table 8

Norfolk Standing Long Jump Scores Compared to Canada Fitness Award Standards (September)

		Above Standard		Below Standard	
Age	Gender	f_o	f_e	f_o	f_e
11	F	37	30.5	24	30.5
11	M	21	27	33	27
12	F	98	104	110	104
12	M	78	94	110	94
13	F	62	79.5	97	79.5
13	M	50	81	112	81
14,15	F	11	9	7	9
14,15	M	11	17.5	24	17.5
All	F	208	223	238	223
All	M	160	219.5	279	219.5

Chi-Square (Females) = (2.02)

Chi-Square (Males) = (32.26), df = 1, p < .001.

females scored below national standards in September, but not significantly. Males scored significantly (.001 level) below standards. With the exception of eleven year olds, all males scored significantly (.05 level) below standard. Thirteen year olds scored significantly below standard at the .001 level. Eleven-year-old females scored best on this test while thirteen-year-old boys were poorest.

Table 9 offers the May results on the standing long jump. Males scored significantly better (.001) than expected, while females scored worse than expected, although not significantly. Only twelve-year-old girls scored as well as expected among the females, while each of the male groups surpassed the expected values. In the standing long jump there was a significant improvement (.001) in performance, largely as a result of a similar improvement by males. Females did not show a significant improvement (Table 10).

While September results for both females and males on the 50-meter dash were below standard, they were not significant, as seen on Table 11. Twelve-year-old females scored best on this test while thirteen-year-old girls were poorest.

As seen on Table 12, Norfolk students scored well on the 50-meter dash in May. Both males and females far surpassed the critical value for the .001 level of significance. Every category but one (fourteen- and fifteen-year-old males) scored significantly better than expected, three (twelve-year-old females, twelve-year-old males, and thirteen-year-old males) at

Table 9

Norfolk Standing Long Jump Scores Compared to Canada Fitness Award Standards (May)

		Above Standard		Below Standard	
Age	Gender	f_o	f_e	f_o	f_e
12	F	107	102	97	102
12	M	129	96	63	96
13	F	69	79.5	90	79.5
13	M	100	80.5	61	80.5
14,15	F	4	9	14	9
14,15	M	19	18.5	18	18.5
All	F	180	190.5	201	190.5
All	M	248	195	142	195

Chi-Square (Females) = (1.16)

Chi-Square (Males) = 28.82, df = 1, $p < .001$.

Table 10

Change in Standing Long Jump Scores from September to May

Overall

	Above Standard		Below Standard	
	f_o	f_e	fo	f_e
September	368	425.40	517	459.60
May	428	370.60	343	400.40

Chi-Square = 32.04, df = 1, $p < .001$.

Males

September	160	216.06	279	222.94
May	248	191.94	142	198.06

Chi-Square = 60.89, df = 1, $p < .001$.

Females

September	208	209.25	238	236.75
May	180	178.75	201	202.25

Chi-Square = 0.04.

Table 11

Norfolk 50-Meter Dash Scores Compared to Canada Fitness Award Standards (September)

		Above Standard		Below Standard	
Age	Gender	f_o	f_e	f_o	f_e
11	F	29	30.5	32	30.5
11	M	31	27	23	27
12	F	121	104	87	104
12	M	86	94	102	94
13	F	61	79.5	98	79.5
13	M	71	81	91	81
14,15	F	7	9	11	9
14,15	M	14	17	21	18
All	F	218	223	228	223
All	M	202	219.5	237	219.5

Chi-Square (Females) = (0.22).

Chi-Square (Males) = (2.79).

Table 12

Norfolk 50-Meter Dash Scores Compared to Canada Fitness Award Standards (May)

		Above Standard		Below Standard	
Age	Gender	f_o	f_e	f_o	f_e
12	F	143	102	61	102
12	M	142	96	50	96
13	F	94	79.5	65	79.5
13	M	112	80.5	49	80.5
14,15	F	15	9	3	9
14,15	M	20	18.5	17	18.5
All	F	252	190.5	129	190.5
All	M	274	195	116	195

Chi-Square (Females) = 39.71, df = 1, $p < .001$.

Chi-Square (Males) = 64.01, df = 1, $p < .001$.

the .001 level. In the 50-meter dash both males and females showed a significant improvement at the .001 level (Table 13).

On curl-ups, seen on Table 14, both females and males scored significantly better (.01 level) than CFA standards in September. 11-year-old boys scored significantly better at the .05 level, while 11-year-old girls, 12-year-old girls, 12-year-old boys, and 13-year-old boys scored significantly better at the .001 level.

May results on the curl-ups test are shown on Table 15. Both genders had significantly better results than expected, however the differences were striking. Female performance was significantly different (.05) largely because of 12 year olds, who were significantly better at the .001 level; 13, 14 and 15 year olds were worse than expected. Each of the male groups surpassed the expected values, with 12 and 13 year olds at the .001 level. In curl-ups there was an overall decrease in performance, though not significant (Table 16). While males scored slightly better than expected, females scored more poorly than expected.

Table 17 shows September performance in push ups. Females scored below standard, but not significantly so. However, the boys scored significantly below standard (.05 level), largely because all male groups scored below standard, especially 13 year olds. Best performance on this test was by 12-year-old females, the worst by 11-year-old females.

Table 18 shows the May results of the push up test. Both males and females scored significantly better than expected (.001

Table 13

Change in 50-Meter Dash Scores from September to May

	Above Standard		Below Standard	
	f_o	f_e	f_o	f_e
Overall				
September	420	505.56	465	379.44
May	526	440.44	245	330.56

Chi-Square = 72.54, df = 1, $p < .001$.

Males				
September	202	252.07	237	186.93
May	274	223.93	116	166.07

Chi-Square = 49.66, df = 1, $p < .001$.

Females				
September	218	253.47	228	192.53
May	252	216.53	129	164.47

Chi-Square = 24.95, df = 1, $p < .001$.

Table 14

Norfolk Curl-up Scores Compared to Canada Fitness Award Standards (September)

		Above Standard		Below Standard	
Age	Gender	f_o	f_e	f_o	f_e
11	F	45	30.5	16	39.5
11	M	36	27	18	27
12	F	134	104	74	104
12	M	123	94	65	94
13	F	74	79.5	85	79.5
13	M	114	81	48	81
14,15	F	12	9	6	9
14,15	M	21	17.5	14	17.5
All	F	265	223	181	223
All	M	294	219.5	145	219.5

Chi-Square (Females) = 15.82, df = 1, $p < .001$.

Chi-Square (Males) = 50.57, df = 1, $p < .001$.

Table 15

Norfolk Curl-up Scores Compared to Canada Fitness Award Standards (May)

		Above Standard		Below Standard	
Age	Gender	f_o	f_e	f_o	f_e
12	F	136	102	68	102
12	M	139	96	53	96
13	F	68	79.5	91	79.5
13	M	112	80.5	49	80.5
14,15	F	7	9	11	9
14,15	M	19	18.5	18	18.5
All	F	211	190.5	170	190.5
All	M	270	195	120	195

Chi-Square (Females) = 4.41, df = 1, $p < .05$.

Chi-Square (Males) = 57.69, df = 1, $p < .001$.

Table 16

Change in Curl-ups Scores from September to May

	Overall			
	Above Standard		Below Standard	
	f_o	f_e	f_o	f_e
September	559	555.80	326	329.20
May	481	484.20	290	286.80

Chi-Square = (0.11).

	Males			
September	294	298.67	145	140.33
May	270	265.33	120	124.67

Chi-Square = 0.48.

	Females			
September	265	256.71	181	189.29
May	211	219.29	170	161.71

Chi-Square = (1.37).

Table 17

Norfolk Push Up Scores Compared to Canada Fitness Award Standards (September)

		Above Standard		Below Standard	
Age	Gender	f_o	f_e	f_o	f_e
11	F	18	30.5	43	30.5
11	M	26	27	28	27
12	F	110	104	98	104
12	M	89	94	99	94
13	F	76	79.5	83	79.5
13	M	67	81	95	81
14,15	F	10	9	8	9
14,15	M	13	17.5	22	17.5
All	F	214	223	232	223
All	M	195	219.5	244	219.5

Chi-Square (Females) = (0.73).

Chi-Square (Males) = (5.25), df = 1, $p < .05$.

Table 18

Norfolk Push Up Scores Compared to Canada Fitness Award Standards (May)

		Above Standard		Below Standard	
Age	Gender	f_o	f_e	f_o	f_e
12	F	137	102	67	102
12	M	134	96	58	96
13	F	88	79.5	71	79.5
13	M	88	80.5	73	80.5
14,15	F	10	9	8	9
14,15	M	22	18.5	15	18.5
All	F	235	190.5	146	190.5
All	M	244	195	146	195

Chi-Square (Females) = 20.79, df = 1, $p < .001$.

Chi-Square (Males) = 24.63, df = 1, $p < .001$.

level). While this is the result of the performances of 12-year-old males and females, every category scored better than expected. As shown in Table 19, in the push up test students showed a significant improvement (.001 level) over the course of the year.

September distance run performances are shown on Table 20. Females scored exactly as expected. Males in general scored significantly below standard (.005 level). Best performance was by 12-year-old girls, worst by 13-year-old boys.

On the distance run in May (Table 21), both males and females were significantly better than expected. Each category did better than expected, especially thirteen-year-old boys (.05) and thirteen-year-old girls (.005). These results indicate a significant improvement for both males (.001 level) and females (.01 level) (Table 22).

On the September shuttle run (Table 23), all groups scored below standard. The male performance was not significantly below, but the females' was, 13-year-old girls being the greatest factor. Top performance on this test was by 11-year-old boys, the worst by 13-year-old girls.

Table 24 shows the May results of the shuttle run. Males and females bettered the expected results significantly (.001). All categories scored better than expected; four scored significantly better. These results indicated a significant improvement at the .001 level for both genders (Table 25).

Table 19

Change in Push Up Scores from September to May

	Overall			
	Above Standard		Below Standard	
	f_o	f_e	f_o	f_e
September	409	474.57	476	410.43
May	479	413.43	292	357.57

Chi-Square = 41.96, df = 1, $p < .001$.

	Males			
September	195	232.47	244	206.53
May	244	206.53	146	183.47

Chi-Square = 27.29, df = 1, $p < .001$.

	Females			
September	214	242.15	232	203.85
May	235	206.85	146	174.15

Chi-Square = 15.54, df = 1, $p < .001$.

Table 20

Norfolk Distance Run Scores Compared to Canada Fitness Award Standards (September)

		Above Standard		Below Standard	
Age	Gender	f_o	f_e	f_o	f_e
11	F	21	30.5	40	30.5
11	M	25	27	29	27
12	F	117	104	91	104
12	M	88	94	100	94
13	F	76	79.5	83	79.5
13	M	61	81	101	81
14,15	F	9	9	9	9
14,15	M	13	17.5	22	17.5
All	F	223	223	223	223
All	M	187	219.5	252	219.5

Chi-Square (Females) = 0.00

Chi-Square (Males) = (9.62), df = 1, $p < .005$.

Table 21

Norfolk Distance Run Scores Compared to Canada Fitness Award Standards (May)

		Above Standard		Below Standard	
Age	Gender	f_o	f_e	f_o	f_e
12	F	115	102	89	102
12	M	104	96	88	96
13	F	99	79.5	60	79.5
13	M	96	80.5	65	80.5
14,15	F	13	9	5	9
14,15	M	23	18.5	14	18.5
All	F	227	190.5	154	190.5
All	M	223	195	167	195

Chi-Square (Females) = 13.99, df = 1, $p < .001$.

Chi-Square (Males) = 8.04, df = 1, $p < .005$.

Table 22

Change in Distance Run Scores from September to May

Overall

	Above Standard		Below Standard	
	f_o	f_e	f_o	f_e
September	410	459.60	475	425.40
May	450	400.40	321	370.60

Chi-Square = 23.91, df = 1, $p < .001$.

Males

September	187	217.12	252	221.88
May	223	192.88	167	197.12

Chi-Square = 17.57, df = 1, $p < .001$.

Females

September	223	242.68	223	203.32
May	227	207.32	154	173.68

Chi-Square = 7.60, df = 1, $p < .01$.

Table 23

Norfolk Shuttle Run Scores Compared to Canada Fitness Award Standards (September)

		Above Standard		Below Standard	
Age	Gender	f_o	f_e	f_o	f_e
11	F	26	30.5	35	30.5
11	M	25	27	29	27
12	F	94	104	114	104
12	M	87	94	101	94
13	F	66	79.5	93	79.5
13	M	76	81	86	81
14,15	F	6	9	12	9
14,15	M	14	17.5	21	17.5
All	F	192	223	254	223
All	M	202	219.5	237	219.5

Chi-Square (Females) = (8.62), df = 1, $p < .005$.

Chi-Square (Males) = (2.79).

Table 24

Norfolk Shuttle Run Scores Compared to Canada Fitness Award Standards (May)

		Above Standard		Below Standard	
Age	Gender	f_o	f_e	f_o	f_e
12	F	126	102	78	102
12	M	133	96	59	96
13	F	108	79.5	51	79.5
13	M	96	80.5	65	80.5
14,15	F	11	9	7	9
14,15	M	21	18.5	16	18.5
All	F	245	190.5	136	190.5
All	M	250	195	140	195

Chi-Square (Females) = 31.18, df = 1, $p < .001$.

Chi-Square (Males) = 31.03, df = 1, $p < .001$.

Table 25

Change in Shuttle Run Scores from September to May

Overall

	Above Standard		Below Standard	
	f_o	f_e	f_o	f_e
September	394	475.10	491	409.90
May	495	413.90	276	357.10

Chi-Square = 64.20, df = 1, p < .001.

Males

September	202	239.36	237	199.64
May	250	212.64	140	177.36

Chi-Square = 27.25, df = 1, p < .001.

Females

September	192	235.67	254	210.33
May	245	201.33	136	179.67

Chi-Square = 37.24, df = 1, p < .001.

Fitness Performance and Age

Table 26 indicates that there was no significant relationship between total fitness and age in September.

As in September, there was no relationship in May between total fitness and age (Table 26).

As can be seen on Table 27, there was a significant relationship (.05 level) between age and performance on three CFA tests. Scores for the standing long jump, 50-meter dash, and distance run exceeded the critical value of 7.82 (df=3). This means that there were significant differences in performance on these tests, according to age.

Table 27 also shows how the May test performance varied with respect to age. For each gender four of the six tests indicated that performance was related to the age of the student.

When gender was also considered (Table 27), the differences were striking. For males, there is a significant relationship between age and performance on only one of the CFA tests. Conversely, for females, there is a significant relationship for all but one of the tests (shuttle run). Three of the tests show significance at the .05 level (standing long jump, push ups, and distance run), and two at the .001 level (50-meter dash and curl-ups).

The significance of age to performance on the standing long jump is due largely to 13 year olds, who scored below standard. Overall, 11 year olds scored best on this test. For females, the significant performance variance was due to the above-standard

Table 26

<u>Fitness and Age (September)</u>

	Above Standard		Below Standard	
Age	f_o	f_e	f_o	f_e
11	47	47.69	68	67.31
12	176	164.22	220	231.78
13	126	133.12	195	187.88
14,15	18	21.98	35	31.02

Chi-Square = 3.35.

<u>May</u>

	Above Standard		Below Standard	
Age	f_o	f_e	f_o	f_e
12	129	121.60	91	98.40
13	196	202.19	168	162.81
14,15	110	112.20	93	90.80

Chi-Square = 1.40.

Table 27

Age and Specific Canada Fitness Award Subtests (September)

	Chi-Square		
CFA Subtest	Females	Males	Overall
Standing Long Jump	10.09***	4.75*	10.97****
50-Meter Dash	14.99****	3.63	11.19****
Curl-ups	18.70****	1.84	6.05*
Push Ups	10.75***	2.31	6.06*
Distance Run	9.48***	3.70	8.77***
Shuttle Run	1.25	0.55	1.23

May

Standing Long Jump	7.68**	3.62	7.79**
50-Meter Dash	7.30**	5.94*	5.28*
Curl-ups	22.73****	6.48*	18.88****
Push Ups	5.57*	8.74***	13.98****
Distance Run	2.53	1.48	3.53
Shuttle Run	1.55	4.50*	1.15

$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .005$. $^{****}p < .001$.

performance of 11 year olds.

On the 50-meter dash, much of the overall performance variance was again caused by the substandard performance of 13 year olds. The significant performance variance for females was created by the above-standard performance of 12 year olds and the substandard performance by thirteen year olds.

For curl-ups, the significance of the female performance variance was due to the poorer performance by 13 year olds.

For push ups, the significance in female performance variance was due to the substandard scores of 11 year olds.

The significance of overall performance variance was largely due to the above-standard performance of 12 year olds. For females, the significance was attributable to substandard performance by 11 year olds.

After comparing the age-specific results among themselves, they were compared with national standards. As can be seen on Table 28, all groups scored below CFA standards in September. 12 year olds and 14-15 year olds were significantly below at the .05 level, while 13 year olds were significantly below at the .001 level. From the September data, it was apparent that 11 year olds were most fit and 13 year olds least fit.

In May, while each of the three age groups (there were no 11 year olds in May) scored better than expected, Table 28 shows that only 12 year olds were significantly better than expected.

Table 29 indicates that the performance of each age group is significantly better in May than in September.

Table 28

Comparison of Norfolk Fitness Scores by Age and Canada Fitness Award Standards (September)

	Above Standard		Below Standard	
Age	f_o	f_e	f_o	f_e
11	47	57.5	68	57.5
12	176	198	220	198
13	126	160.5	195	160.5
14,15	18	26	35	27

Chi-Square = 29.00, df = 1, $p < .001$.

May

	Above Standard		Below Standard	
Age	f_o	f_e	f_o	f_e
12	129	110	91	110
13	196	182	168	182
14,15	110	101.5	93	101.5

Chi-Square = 10.13, df = 1, $p < .005$.

Table 29

Change in Fitness by Age from September to May

12 Year Olds

	Above Standard		Below Standard	
	f_o	f_e	f_o	f_e
September	176	196.07	220	199.93
May	129	108.93	91	111.07

Chi-Square = 11.39, df = 1, $p < .001$.

13 Year Olds

September	126	150.89	195	170.11
May	196	171.11	168	192.89

Chi-Square = 14.58. df =1, $p < .001$.

14 Year Olds

September	18	26.50	35	26.50
May	110	101.50	93	101.50

Chi-Square = 6.88, df = 1, $p < .01$,

Fitness Performance and Grade

As can be seen on Table 30, there was no significant relationship in September between fitness and grade. The Chi-Square values for each gender, and overall, fall short of the critical value of 3.84 (df=1).

As can be seen in Table 31, results in May again indicate that there was no relationship between grade and fitness level for boys nor girls.

After comparing Norfolk scores for grades seven and eight students, these scores were compared with CFA standards (Table 32).

Both grade seven scores and grade eight scores were significantly (.001 level) poorer than expected. Contributing to this significance were grade seven boys (.001 level) and grade eight girls (.001 level).

Table 32 shows that Norfolk grades seven and eight students were fit in May. Although both genders in both grades surpassed the expected values, only grade 8 boys scored significantly better than expected.

When May and September results are compared (Tables 33, 34), it is apparent that there was a significant fitness improvement in both genders for both grades.

Fitness Performance and School

As can be seen on Table 35, in September there was a significant difference in fitness by school. In May there was

Table 30

Fitness and Grade (September)

	Overall			
	Grade 7		Grade 8	
	f_o	f_e	f_o	f_e
Above Standard	197	196.15	170	170.85
Below Standard	276	276.85	242	241.15

Chi-Square = 0.00.

	Females			
Above Standard	104	96.42	79	86.58
Below Standard	131	138.58	132	124.42

Chi-Square = 2.13.

	Males			
Above Standard	93	99.75	91	84.25
Below Standard	145	138.25	110	116.75

Chi-Square = 1.72.

Table 31

Fitness and Grade (May)

Overall

	Grade 7		Grade 8	
	f_o	f_e	f_o	f_e
Above Standard	238	240.99	197	194.01
Below Standard	198	195.01	154	156.99

Chi-Square = 0.20.

Females

Above Standard	117	114.15	90	92.85
Below Standard	92	94.85	80	77.15

Chi-Square = (0.36).

Males

Above Standard	121	126.85	107	101.15
Below Standard	106	100.15	74	79.85

Chi-Square = 1.38.

Table 32

Comparison of Norfolk Fitness Scores by Grade and Canada Fitness Award Standards (September)

		Above Standard		Below Standard	
Gr.	Gender	f_o	f_e	f_o	f_e
7	F	104	117.5	131	117.5
7	M	93	119	145	119
8	F	79	105.5	132	105.5
8	M	91	100.5	110	100.5
7	F,M	197	236.5	276	236.5
8	F,M	170	206	242	206

Chi-Square (Grade 7) = (13.19), df = 1, $p < .001$.

Chi-Square (Grade 8) = (12.58), df = 1, $p < .001$.

May

7	F	117	104.5	92	104.5
7	M	121	113.5	106	113.5
8	F	90	85	80	85
8	M	107	90.5	74	90.5
7	F,M	238	218	198	218
8	F,M	197	175.5	154	175.5

Chi-Square (Grade 7) = 3.67; (Grade 8) = 5.27, df = 1, $p < .05$.

Table 33

Fitness Change for Grade Sevens from September to May

	Overall			
	Above Standard		Below Standard	
	f_o	f_e	f_o	f_e
September	197	226.35	276	246.65
May	238	208.65	198	227.35

Chi-Square = 15.22, df = 1, $p < .001$.

	Females			
September	104	116.97	131	118.03
May	117	104.03	92	104.98

Chi-Square = 6.09, df = 1, $p < .05$.

	Males			
September	93	109.53	145	128.47
May	121	104.47	106	122.53

Chi-Square = 9.47, df = 1, $p < .005$.

Table 34

Fitness Change for Grade Eights from September to May

	Overall			
	Above Standard		Below Standard	
	f_o	f_e	f_o	f_e
September	170	198.17	242	213.83
May	197	168.83	154	182.17

Chi-Square = 16.77, df = 1, $p < .001$.

	Females			
September	79	93.59	132	117.41
May	90	75.41	80	94.59

Chi-Square = 9.15, df = 1, $p < .005$.

	Males			
September	91	104.18	110	96.82
May	107	93.82	74	87.18

Chi-Square = 7.30, df = 1, $p < .01$.

Table 35

Fitness of All Students by School (September)

	Above Standard		Below Standard	
School	f_o	f_e	f_o	f_e
1	5	9.12	17	12.88
2	24	24.88	36	35.12
3	26	15.76	12	22.24
4	31	34.83	53	49.17
5	68	66.76	93	94.24
6	95	99.94	146	141.06
7	3	13.27	29	18.73
8	11	14.10	23	19.90
9	6	7.46	12	10.54
10	11	20.73	39	29.27
11	59	37.32	31	52.68
12	28	22.81	27	32.19

Chi-Square = 62.32, df = 1, $p < .001$.

again a significant relationship between these two variables (Table 36).

Contributing to the significant overall Chi-Square value in September were Schools 3, 7, 10, and 11. Schools 3 and 11 performed better than expected, while Schools 7 and 10 did not score as well as expected. In May the performance variance was largely due to the scores of schools 5, 10, and 11. School 10 scored more poorly than expected, while schools 5 and 11 scored better than expected.

With respect to male scores (Table 37), there were better than expected values for School 3 and, especially, School 11 and poor results for School 7. These contributed greatly to the significant relationship between male fitness scores and school. In May, this same significant relationship existed (Chart 38), largely because of schools 5, 10, and 11. School 10 scored more poorly than expected, and 5 and 11 scored better than expected.

When considering females only in September (Table 39), there were poorer than expected performances by Schools 7, 8, and 10, along with the better than expected performances by Schools 3 and 11. These contributed to a significant relationship between the two variables. In May, the same relationship existed (Table 40), largely due to the poor performance of school 10.

After comparing the scores of Norfolk schools, these were compared with CFA standards (Table 41). In September, nine of the twelve schools had results which fell below standard. Three scored significantly more poorly at the .05 level (Schools 1, 4,

Table 36

Fitness of All Students by School (May)

	Above Standard		Below Standard	
School	f_o	f_e	f_o	f_e
1	7	11.61	14	9.39
2	31	31.51	26	25.49
3	14	13.82	11	11.18
4	41	39.24	30	31.76
5	101	84.02	51	67.98
6	113	123.26	110	99.74
7	11	16.58	19	13.42
8	15	12.16	7	9.84
9	6	6.08	5	4.92
10	8	25.98	39	21.02
11	61	43.67	18	35.33
12	27	27.08	22	21.92

Chi-Square = 62.75, df = 1, $p < .001$.

Table 37

Fitness of Males by School (September)

	Above Standard		Below Standard	
School	f_o	f_e	f_o	f_e
1	2	4.19	8	5.81
2	11	12.15	18	16.85
3	14	8.38	6	11.62
4	17	18.44	27	25.56
5	27	33.53	53	46.47
6	48	48.62	68	67.38
7	0	6.29	15	8.71
8	9	7.96	10	11.04
9	3	3.77	6	5.23
10	5	7.54	13	10.46
11	35	21.79	17	30.21
12	13	11.32	14	15.68

Chi-Square = 38.05, df = 1, $p < .001$.

Table 38

Fitness of Males by School (May)

	Above Standard		Below Standard	
School	f_o	f_e	f_o	f_e
1	3	5.59	7	4.41
2	10	15.65	18	12.35
3	8	7.26	5	5.74
4	24	23.47	18	18.53
5	53	42.47	23	33.53
6	57	62.59	55	49.41
7	4	7.82	10	6.18
8	10	8.94	6	7.06
9	2	3.35	4	2.65
10	5	10.06	13	7.94
11	37	26.26	10	20.74
12	15	14.53	11	11.47

Chi-Square = 36.11, df = 1, $p < .001$.

Table 39

Fitness of Females by School (September)

	Above Standard		Below Standard	
School	f_o	f_e	f_o	f_e
1	3	4.92	9	7.08
2	13	12.72	18	18.28
3	12	7.39	6	10.61
4	14	16.41	26	23.59
5	41	33.24	40	47.76
6	47	51.29	78	73.71
7	3	6.98	14	10.02
8	2	6.15	13	8.85
9	3	3.69	6	5.31
10	6	13.13	26	18.87
11	24	15.59	14	22.41
12	15	11.49	13	16.51

Chi-Square = 35.34, df = 1, $p < .001$.

Table 40

Fitness of Females by School (May)

	Above Standard		Below Standard	
School	f_o	f_e	f_o	f_e
1	4	5.94	7	5.06
2	16	12.96	8	11.04
3	6	6.48	6	5.52
4	17	15.66	12	13.34
5	48	41.05	28	34.95
6	56	59.95	55	51.05
7	7	8.64	9	7.36
8	5	3.24	1	2.76
9	4	2.70	1	2.30
10	3	15.66	26	13.34
11	24	17.28	8	14.72
12	12	12.42	11	10.58

Chi-Square = 38.44, df = 1, $p < .001$.

Table 41

Comparison of Norfolk Fitness Scores by School and Canada Fitness Award Standards (September)

	Above Standard		Below Standard	
School	f_o	f_e	f_o	f_e
1	5	11	17	11
2	24	30	36	30
3	26	19	12	19
4	31	42	53	42
5	68	80.5	93	80.5
6	95	120.5	146	120.5
7	3	16	29	16
8	11	17	23	17
9	6	9	12	9
10	11	25	39	25
11	59	45	31	45
12	28	27.5	27	27.5

Chi-Square = (86.32), df = 1, $p < .001$.

and 8), while another scored significantly more poorly at the .005 level (School 6), and two others more poorly at the .001 level (Schools 7 and 10). Two schools scored significantly better than standards: Schools 3 (.05 level) and 11 (.005 level).

Table 42 shows the fitness results for each school in May. While nine of 12 schools scored better than expected, only two schools (11 and 5) scored significantly better (.001 level). One school, #10, scored significantly more poorly than expected (.001 level).

Table 43 shows the fitness change within Norfolk schools. All but one school, 3, scored as well as expected. In addition, nine schools had scores significantly higher than expected, with #5 improving the most. It is clear that there was a significant relationship between fitness change and school (Table 44).

Gender differences exist in school scores. While eight schools had male scores below CFA standards in September (Table 45), only two were significantly poor: Schools 5 (.005 level) and 7 (.001 level). One school (11) was significantly better than expected (.05 level). Table 46 shows the May fitness results of boys for each school. Boys at two schools, 11 and 5, were significantly more fit than expected (.001 level). Five other schools scored better than expected but not significantly so.

Table 47 shows the fitness change for boys within each school. Each school scored better than expected, with eight

Table 42

Comparison of Norfolk Fitness Scores by School and Canada Fitness Award Standards (May)

	Above Standard		Below Standard	
School	f_o	f_e	f_o	f_e
1	7	10.5	14	10.5
2	31	28.5	26	28.5
3	14	12.5	11	12.5
4	41	35.5	30	35.5
5	101	76	51	76
6	113	111.5	110	111.5
7	11	15	19	15
8	15	11	7	11
9	6	5.5	5	6
10	8	23.5	39	23.5
11	61	39.5	18	39.5
12	27	24.5	22	24.5

Chi-Square = 70.57, df = 1, $p < .001$.

Table 43

Fitness Change by School from September to May

	Improved		Not Improved	
School	f_o	f_e	f_o	f_e
1	14	10.5	7	10.5
2	42	28.5	15	28.5
3	10	12.5	15	12.5
4	61	35.5	10	35.5
5	125	76	27	76
6	166	111.5	57	111.5
7	28	15	2	15
8	19	10.5	2	10.5
9	7	5.5	4	5.5
10	24	23.5	23	23.5
11	66	39.5	13	39.5
12	33	24	15	24

Chi-Square = 248.65, df = 1, $p < .001$.

Table 44

Fitness Change and School

	Improved		Not Improved	
School	f_o	f_e	f_o	f_e
1	14	15.92	7	5.08
2	42	43.20	15	13.80
3	10	18.95	15	6.06
4	61	53.82	10	17.18
5	125	115.21	27	36.79
6	166	169.03	57	53.97
7	28	22.74	2	7.26
8	19	15.92	2	5.08
9	7	8.34	4	2.66
10	24	35.62	23	11.38
11	66	59.88	13	19.12
12	33	36.38	15	11.62

Chi-Square = 54.10, df = 1, p < .001.

Table 45

Comparison of Norfolk Boys' Fitness Scores by School and Canada Fitness Award Standards (September)

	Above Standard		Below Standard	
School	f_o	f_e	f_o	f_e
1	2	5	8	5
2	11	14.5	18	14.5
3	14	10	6	10
4	17	22	27	22
5	27	40	53	40
6	48	58	68	58
7	0	7.5	15	7.5
8	9	9.5	10	9.5
9	3	4.5	6	4.5
10	5	9	13	9
11	35	26	17	26
12	13	13.5	14	13.5

Chi-Square = (11.48), df = 1, $p < .001$.

Table 46

Comparison of Norfolk Boys' Fitness Scores by School and Canada Fitness Award Standards (May)

	Above Standard		Below Standard	
School	f_o	f_e	f_o	f_e
1	3	5	7	5
2	10	14	18	14
3	8	6.5	5	6.5
4	24	21	18	21
5	53	38	23	38
6	57	56	55	56
7	4	7	10	7
8	10	8	6	8
9	2	3	4	3
10	5	9	13	9
11	37	23.5	10	23.5
12	15	13	11	13

Chi-Square = 41.25, df = 1, $p < .001$.

Table 47

Change in Boys' Fitness Scores by School from September to May

	Improved		Not Improved	
School	f_o	f_e	f_o	f_e
1	8	5	2	5
2	18	14	10	14
3	8	6.5	5	6.5
4	36	21	6	21
5	71	38	5	38
6	87	56	25	56
7	13	7	1	7
8	14	7.5	1	7.5
9	4	3	2	3
10	16	9	2	9
11	40	23.5	7	23.5
12	19	13	7	13

Chi-Square = 181.48, df = 1, $p < .001$.

schools experiencing significant improvement. For females, eight of twelve schools also scored below standards in September (Table 48). One school scored significantly more poorly at the .05 level (School 7), two others at the .01 level (Schools 6 and 8), and one at the .001 level (School 10).

Table 49 shows the school results for females in May. Again, 9 of 12 schools scored better than expected, but in this case three schools were significantly better: 11 (.005 level), 2 (.05 level), and 5 (.05 level). One school, 10, scored significantly more poorly than expected (.001 level).

Table 50 shows the fitness change for girls within each school. The results are similar to those of the boys: 10 of the 12 schools scored better than expected. However, two schools experienced a significant decrease in performance: 3 (.005 level) and 10 (.01 level).

Fitness Performance and Extracurricular Physical Activity

Table 51 shows the relationship between fitness performance in September and participation in regular physical activity during the summer (twice weekly).

Again in May, there was a significant relationship (.001) between total fitness and extracurricular physical activity for both males and females (Table 52). Contrary to September, however, extracurricular activity was more important for females.

It is evident that there existed a significant positive relationship between the performance of students in September and

Table 48

Comparison of Norfolk Girls' Fitness Scores by School and Canada Fitness Award Standards (September)

	Above Standard		Below Standard	
School	f_o	f_e	f_o	f_e
1	3	6	9	6
2	13	15.5	18	15.5
3	12	9	6	9
4	14	20	26	20
5	41	40.5	40	40.5
6	47	62.5	78	62.5
7	3	8.5	14	8.5
8	2	7.5	13	7.5
9	3	4.5	6	4.5
10	6	16	26	16
11	24	19	14	19
12	15	14	13	14

Chi-Square = (48.57), df = 1, $p < .001$.

Table 49

Comparison of Norfolk Girls' Fitness Scores by School and Canada Fitness Award Standards (May)

	Above Standard		Below Standard	
School	f_o	f_e	f_o	f_e
1	4	5.5	7	5.5
2	21	14.5	8	14.5
3	6	6	6	6
4	17	14.5	12	14.5
5	48	38	28	38
6	56	55.5	55	55.5
7	7	8	9	8
8	5	3	1	3
9	4	2.5	1	2.5
10	3	14.5	26	14.5
11	24	16	8	16
12	12	11.5	11	11.5

Chi-Square = 43.78, df = 1, $p < .001$.

Table 50

Change in Girls' Fitness Scores by School from September to May

	Improved		Not Improved	
School	f_o	f_e	f_o	f_e
1	6	5.5	5	5.5
2	24	14.5	5	14.5
3	2	6	10	6
4	25	14.5	4	14.5
5	54	38	22	38
6	79	55.5	32	55.5
7	15	8	1	8
8	5	3	1	3
9	3	2.5	2	2.5
10	8	14.5	21	14.5
11	26	16	6	16
12	14	11	8	11

Chi-Square = 101.54, df = 1, $p < .001$.

Table 51

Fitness and Extracurricular Physical Activity (September)

Overall

	Extra Activity		No Activity	
	f_o	f_e	f_o	f_e
Above Standard	189	148.04	178	218.96
Below Standard	168	208.96	350	309.04

Chi-Square = 32.45, df = 1, p < .001.

Females

Above Standard	79	65.24	104	117.76
Below Standard	80	93.76	183	169.24

Chi-Square = 7.65, df = 1, p < .01.

Males

Above Standard	110	82.99	74	101.01
Below Standard	88	115.01	167	139.99

Chi-Square = 27.56, df = 1, p < .001.

Table 52

Fitness and Extracurricular Physical Activity (May)

	Overall			
	Extra Activity		No Activity	
	f_o	f_e	f_o	f_e
Above Standard	332	287.02	67	111.98
Below Standard	196	240.98	139	94.02

Chi-Square = 55.04, df = 1, $p < .001$.

	Females			
Above Standard	159	133.93	37	62.07
Below Standard	87	112.07	77	51.93

Chi-Square = 32.53, df = 1, $p < .001$.

	Males			
Above Standard	174	153.95	30	50.05
Below Standard	109	129.05	62	41.95

Chi-Square = 23.34, df = 1, $p < .001$.

physical activity carried on during the summer. Overall, the relationship was significant at the .001 level of confidence. When considering gender, it is shown that, while summer activity was significantly important (.01 level) in predicting performance for females, it was much more a factor for males (.001 level). When considering age and gender (Table 53), the pattern is more distinct. For females, only one age group (12 year olds) showed a significant relationship. For males, there is only one group (11 year olds) for whom there was not a significant relationship.

Table 54 shows that in May extracurricular activity was significantly related to total fitness for every age and gender category, with the exception of 14 and 15 year old boys.

After comparing participating and non-participating students within Norfolk, these results were also compared with CFA standards. As can be seen on Table 55, in September those who had participated in extramural physical activity surpassed CFA standards. More significant than the results of those who participated were the results of the non-participants: those students scored far below standard, contributing greatly to the significantly different (.001 level) results than were expected.

Table 56 indicates how participating and non-participating students performed in May in comparison with CFA standards. For both males and females, participating students scored significantly better than expected, while non-participating students scored significantly more poorly.

Table 57 shows the relationship between fitness change and

Table 53

Fitness by Age and Extracurricular Physical Activity (September)

Age	Gender	Chi-Square
11	F	0.97
11	M	1.16
12	F	5.42*
12	M	17.97****
13	F	1.50
13	M	4.35*
14,15	F	0.00
14,15	M	8.52***

*$p < .05$. ***$p < .005$. ****$p < .001$.

Table 54

Fitness and Extracurricular Physical Activity (May)

Age	Gender	Chi-Square
12	F	12.27****
12	M	9.78***
13	F	16.14****
13	M	15.59****
14,15	F	5.76*
14,15	M	1.27

*p< .05. ***p< .005. ****p< .001.

Table 55

Comparison of Norfolk Fitness Scores by Extracurricular Physical Activity and Canada Fitness Award Standards (September)

	Overall			
	Above Standard		Below Standard	
	f_o	f_e	f_o	f_e
Extra Activity	189	178.5	168	178.5
No Activity	178	264	350	264

Chi-Square = 57.27, df = 1, $p < .001$.

	Females			
Extra Activity	79	79.5	80	79.5
No Activity	104	143.5	183	143.5

Chi-Square = 21.76, df = 1, $p < .001$.

	Males			
Extra Activity	110	99	88	99
No Activity	74	120.5	167	120.5

Chi-Square = 38.33, df = 1, $p < .001$.

Table 56

Comparison of Norfolk Fitness Scores by Extracurricular Physical Activity and Canada Fitness Award Standards (May)

Overall

	Above Standard		Below Standard	
	f_o	f_e	f_o	f_e
Extra Activity	333	264.5	196	264.5
No Activity	67	103	139	103

Chi-Square = 60.65, df = 1, $p < .001$.

Females

Extra Activity	159	123	87	123
No Activity	37	57	77	57

Chi-Square = 35.11, df = 1, $p < .001$.

Males

Extra Activity	174	141.5	109	141.5
No Activity	30	46	62	46

Chi-Square= 26.23, df = 1, $p < .001$.

Table 57

Fitness Change and Extracurricular Physical Activity

	Overall			
	Improvement		No Improvement	
	f_o	f_e	f_o	f_e
Extra Activity	418	396.50	109	130.50
No Activity	135	156.50	73	51.50

Chi-Square = 16.64, df = 1, $p < .001$.

	Females			
Extra Activity	178	168.09	66	75.91
No Activity	70	79.91	46	36.09

Chi-Square = 5.82, df = 1, $p < .05$.

	Males			
Extra Activity	240	230.17	43	52.83
No Activity	65	74.83	27	17.17

Chi-Square = 9.17, df = 1, $p < .005$.

extracurricular physical activity. There was a significant relationship between these variables for females (.005 level) and males (.05 level), and overall (.001 level).

Raw Scores (by Gender)

One of the objectives of this research was to compare the absolute performances of males and females. Up to this point in the presentation of the September results, aggregate scores have been used for means of comparison of Norfolk and national standards. These scores were generated through comparison with standards which acknowledge performance differences by gender and age. At this juncture raw scores will be used in order to address the matter of male-versus-female fitness performance. *t*-tests, which measure the significance of a difference in means, were used to accomplish this task.

When considering eleven year olds (Table 58), boys outperformed girls on every test but the long jump. Boys were significantly better than girls on four of the tests: 50-meter dash (.05), distance run (.005), shuttle run (.005), and push ups (.001).

With twelve year olds (Table 59), the boys outperformed the girls on all six tests in September. Their results were significantly better on four of the tests: curl-ups (.05), push ups (.001), distance run (.001), and shuttle run (.001). In May, the boys outperformed the girls significantly on all tests (Table 60).

Table 58

Gender Comparison for 11 Year Olds (September)

	Females	Males	
CFA Subtest	Mean (N=61)	Mean (N=54)	T-Value
Standing Long Jump	152.01(S=20.03)	149.48(S=23.44)	0.63
50-Meter Dash	9.47(S=0.86)	9.10(S=0.78)	2.37*
Curl-ups	35.95(S=18.92)	40.37(S=25.29)	-1.00
Push Ups	9.46(S=6.34)	16.17(S=10.03)	-4.19****
Distance Run	663.26(S=144.85)	581.02(S=120.76)	-3.18***
Shuttle Run	13.76(S=1.26)	12.99(S=1.21)	-3.25***

*$p < .05$. ***$p < .005$. ****$p < .001$.

Table 59

Gender Comparison for 12 Year Olds (September)

	Females	Males	
CFA Subtest	Mean (N=208)	Mean (N=188)	T-Value
Standing Long Jump	155.07(S=21.95)	158.05(S=25.18)	-1.25
50-Meter Dash	9.28(S=1.09)	9.10(S=1.16)	1.79
Curl-ups	40.39(S=33.98)	45.92(S=25.69)	-2.09*
Push Ups	11.30(S=8.59)	18.69(S=12.99)	-7.19****
Distance Run	648.33(S=146.61)	583.39(S=147.25)	4.11****
Shuttle Run	13.50(S=1.20)	12.90(S=1.43)	3.71****

*$p < .05$. ****$p < .001$.

Table 60

Gender Comparison for 12 Year Olds (May)

	Females	Males	
CFA Subtest	Mean (N=181)	Mean (N=182)	T-Value
Standing Long Jump	161.95(S=21.95)	177.18(S=25.18)	-6.14****
50-Meter Dash	9.04(S=1.09)	8.52(S=1.16)	4.47****
Curl-ups	52.03(S=33.98)	68.09(S=65.69)	-2.93***
Push Ups	16.81(S=8.59)	27.30(S=12.99)	-9.09****
Distance Run	655.79(S=146.61)	558.51(S=147.25)	6.31****
Shuttle Run	13.10(S=1.20)	12.40(S=1.43)	5.07****

***$p < .005$. ****$p < .001$.

Thirteen-year-old boys also surpassed the girls on all six tests (Table 61) in September. The results on all six tests were significantly better (.005 and .001). Again in May, the boys outscored the girls significantly (.001) on all tests (Table 62).

Likewise, fourteen-year-old boys outperformed their female counterparts on all six tests (Table 63) in September. All of the tests showed significant results: curl-ups and 50-meter dash (.05), shuttle run (.01), push ups (.005), distance run (.005), and standing long jump (.001). In May the boys outscored the girls on all tests, but the results were not significant for curl-ups and the 50-meter dash (Table 64).

Raw Scores (by Age)

Another point to consider was the performance of each age group in terms of raw scores.

In September, twelve-year-old females outperformed eleven year olds on all six CFA tests, but not significantly so (Tables 65, 66, 67, 68, 69, and 70). Twelve-year-old boys bettered eleven year olds on five of six tests (Tables 65, 66, 67, 68, 69, and 70), significantly so (.05) on the standing long jump.

For the comparison between thirteen year olds and twelve year olds, the older girls outscored the younger on all tests, but not significantly so (Tables 65, 66, 67, 68, 69, and 70). (It should be noted that the scores on the distance run cannot be compared because the distances are not the same.) Thirteen-year-old boys also outperformed the twelve year olds but, contrary to

Table 61

Gender Comparison for 13 Year Olds (September)

	Females	Males	
CFA Subtest	Mean (N=159)	Mean (N=162)	T-Value
Standing Long Jump	157.55(S=21.99)	168.99(S=25.37)	-4.05****
50-Meter Dash	9.16(S=1.25)	8.74(S=0.98)	3.08***
Curl-ups	40.84(S=20.50)	61.56(S=22.29)	-5.34****
Push Ups	12.01(S=8.16)	21.65(S=13.45)	-8.33****
Distance Run	993.11(S=164.13)	848.41(S=151.44)	7.30****
Shuttle Run	13.45(S=1.61)	12.33(S=1.20)	7.73****

***$p < .005$. ****$p < .001$.

Table 62

Gender Comparison for 13 Year Olds (May)

	Females	Males	
CFA Subtest	Mean (N=130)	Mean (N=144)	T-Value
Standing Long Jump	163.53(S=21.99)	188.17(S=25.37)	-8.61****
50-Meter Dash	8.87(S=1.25)	8.26(S=0.98)	4.51****
Curl-ups	41.62(S=20.50)	67.74(S=62.29)	-4.76****
Push Ups	14.67(S=8.16)	28.26(S=13.45)	-10.22****
Distance Run	957.25(S=164.13)	780.66(S=151.44)	9.22****
Shuttle Run	12.90(S=1.61)	12.09(S=1.20)	4.63****

****$p < .001$.

Table 63

Gender Comparison for 14 and 15 Year Olds (September)

	Females	Males	
CFA Subtest	Mean (N=18)	Mean (N=35)	T-Value
Standing Long Jump	148.83(S=17.01)	176.51(S=31.21)	-3.89****
50-Meter Dash	8.94(S=0.79)	8.44(S=0.92)	2.30*
Curl-ups	38.17(S=24.01)	55.63(S=48.20)	-2.05*
Push Ups	12.33(S=4.63)	21.49(S=12.96)	-3.33***
Distance Run	969.89(S=168.32)	820.86(S=169.19)	3.14***
Shuttle Run	13.52(S=1.98)	12.23(S=1.60)	2.45*

*$p < .05$. ***$p < .005$. ****$p < .001$.

Table 64

Gender Comparison for 14 and 15 Year Olds (May)

	Females	Males	
CFA Subtest	Mean (N=11)	Mean (N=30)	T-Value
Standing Long Jump	159.64(S=17.01)	200.00(S=31.21)	-5.27****
50-Meter Dash	8.57(S=0.79)	8.07(S=0.92)	1.75
Curl-ups	38.91(S=24.01)	57.20(S=48.20)	-1.61
Push Ups	15.73(S=4.63)	30.40(S=12.96)	-5.34****
Distance Run	947.09(S=168.32)	739.67(S=169.19)	3.49***
Shuttle Run	13.39(S=1.98)	12.00(S=1.60)	2.09*

*$p < .05$. ***$p < .005$. ****$p < .001$.

Table 65

Age Comparison for the Standing Long Jump (September)

Females

Age	N	Mean Score	T-Value (with previous age)
11	61	152.10(S=20.03)	-----
12	208	155.07(S=21.95)	-0.99
13	159	157.55(S=21.99)	-1.06
14,15	18	148.83(S=17.01)	1.66

Males

Age	N	Mean Score	T-Value (with previous age)
11	54	149.48(S=23.44)	-----
12	188	158.05(S=25.81)	-2.29*
13	162	168.99(S=25.37)	-3.83**
14,15	35	176.51(S=31.21)	-1.35

$*p < .05$. $**p < .01$.

Table 66

Age Comparison for the Standing Long Jump (May)

Females			
Age	N	Mean Score	T-Value (with previous age)
12	181	161.95(S=21.95)	-----
13	130	163.53(S=21.99)	-0.63
14,15	11	159.64(S=17.01)	0.71
Males			
Age	N	Mean Score	T-Value (with previous age)
12	182	177.18(S=25.18)	-----
13	144	188.17(S=25.37)	-3.90****
14,15	30	200.00(S=31.21)	-1.95

****$p<$ .001.

Table 67

Age Comparison for the 50-Meter Dash (September)

Females			
Age	N	Mean Score	T-Value (with previous age)
11	61	9.47(S=0.86)	-----
12	208	9.28(S=1.09)	1.42
13	159	9.16(S=1.25)	1.07
14,15	18	8.94(S=0.79)	1.39

Males			
Age	N	Mean Score	T-Value (with previous age)
11	54	9.10(S=0.78)	-----
12	188	9.10(S=1.16)	0.01
13	162	8.74(S=0.98)	2.90**
14,15	35	8.44(S=0.92)	1.46

**$p < .01$.

Table 68

Age Comparison for the 50-Meter Dash (May)

Females			
Age	N	Mean Score	T-Value (with previous age)
12	181	9.04(S=1.09)	-----
13	130	8.87(S=1.25)	1.24
14,15	11	8.57(S=0.79)	1.16
Males			
Age	N	Mean Score	T-Value (with previous age)
12	182	8.52(S=1.16)	-----
13	144	8.26(S=0.98)	2.17*
14,15	30	8.07(S=0.92)	1.03

*$p < .05$.

Table 69

Age Comparison for Curl-ups (September)

Females			
Age	N	Mean Score	T-Value (with previous age)
11	62	35.95(S=18.92)	-----
12	208	40.39(S=33.98)	-1.49
13	159	40.84(S=20.50)	-0.17
14,15	18	38.17(S=24.01)	0.61

Males			
Age	N	Mean Score	T-Value (with previous age)
11	54	40.37(S=25.29)	-----
12	188	45.92(S=25.69)	-1.32
13	162	61.56(S=22.29)	-3.99**
14,15	35	55.63(S=48.20)	0.72

$^{**}p < .01$.

Table 70

Age Comparison for Curl-ups (May)

Females			
Age	N	Mean Score	T-Value (with previous age)
12	181	52.03(S=33.98)	-----
13	130	41.62(S=20.50)	3.36****
14,15	11	38.91(S=24.01)	0.36

Males			
Age	N	Mean Score	T-Value (with previous age)
12	182	68.09(S=65.69)	-----
13	144	67.74(S=62.29)	0.05
14,15	30	57.20(S=48.20)	1.03

****$p < .001$.

the girls, all of the results were significantly better (Tables 65, 66, 67, 68, 69, and 70): push ups (.05), standing long jump (.01), 50-meter dash (.01), curl-ups (.01), and shuttle run (.01).

For girls, fourteen year olds scored more poorly than thirteen year olds on four of the tests, while outscoring them on two. The results were not significant, however (Tables 65, 66, 67, 68, 69, and 70). Fourteen-year-old boys surpassed thirteen year olds on four of six tests, but the results were not significant (Tables 65, 66, 67, 68, 69, and 70).

For 12- and 13-year-old females, there was not a clear picture in September as to physical superiority. While 13 year olds surpassed 12 year olds on three CFA subtests, the scores were not significantly different. In addition, while 12 year olds only surpassed 13 year olds on two tests, there were significant differences in the scores on those two tests.

When 13- and 14-year-old girls were compared, the same situation existed. Thirteen year olds surpassed 14 year olds on three subtests, while 14 year olds were superior on the remaining three. In no subtest were the scores of 13 and 14 year olds significantly different.

For boys there was a more definitive pattern in September. When comparing 12 and 13 year olds, 12 year olds bettered 13 year olds on only one test and the results were very close. On the remaining four subtests 13 year olds significantly bettered the 12 year olds.

Fourteen-year-old boys surpassed 13 year olds on five of six tests, but with results which were not significantly different. Interestingly, the one subtest which differed from the other five when comparing 12 and 13 year olds, and 13 and 14 year olds, was the same: curl-ups.

In May, for males, every subtest indicated the physical superiority of any age over the previous one. It should be stated, however, that in most cases the differences were not significant (Tables 71, 72, 73, 74, 75, and 76). For females, there was no pattern in the comparison of age-specific results (Tables 71, 72, 73, 74, 75, and 76).

Initial Fitness Level and Fitness Change

One of the questions to be addressed within this research was whether or not physical eduction instruction proved more beneficial to any particular group of students, in terms of physical abilities. Table 77 indicates that there was no significant relationship between the initial fitness level of the student and the fitness change which students exhibited. This holds true for both females and males. Interestingly, however, there was a trend present: for males and females there was more substantial improvement by those students whose initial fitness level was below the 50th percentile.

Since previous research had suggested that those students below the 50th percentile could be most affected by instruction, the results were divided with respect to the qualifications of

Table 71

Age Comparison for Push Ups (September)

Females			
Age	N	Mean Score	T-Value (with previous age)
11	61	9.46(S=6.34)	-----
12	208	11.30(S=8.59)	-1.93
13	159	12.01(S=8.16)	-0.90
14,15	18	12.33(S=4.63)	0.20

Males			
Age	N	Mean Score	T-Value (with previous age)
11	54	16.17(S=10.03)	-----
12	188	18.69(S=12.99)	-1.53
13	162	21.65(S=13.45)	-2.22*
14,15	35	21.49(S=12.96)	0.07

*$p < .05$.

Table 72

Age Comparison for Push Ups (May)

		Females	
Age	N	Mean Score	T-Value (with previous age)
12	181	16.81(S=8.59)	-----
13	130	14.67(S=8.16)	2.23*
14,15	11	15.73(S=4.63)	-0.67

		Males	
Age	N	Mean Score	T-Value (with previous age)
12	182	27.30(S=12.99)	-----
13	144	28.26(S=13.45)	-0.65
14,15	30	30.40(S=12.96)	-0.82

*$p < .05$.

Table 73

Age Comparison for the Distance Run (September)

Females			
Age	N	Mean Score	T-Value (with previous age)
11	61	663.26(S=144.85)	-----
12	208	648.33(S=146.61)	0.65
13	159	993.11(S=164.13)	-----
14,15	18	969.89(S=168.32)	0.64

Males			
Age	N	Mean Score	T-Value (with previous age)
11	54	581.02(S=120.76)	-----
12	188	583.39(S=147.25)	-0.12
13	162	848.41(S=151.44)	-----
14,15	35	820.86(S=169.19)	0.75

Table 74

Age Comparison for the Distance Run (May)

Females

Age	N	Mean Score	T-Value (with previous age)
12	181	655.79(S=146.61)	-----
13	130	957.25(S=164.13)	-----
14,15	11	947.09(S=168.32)	0.19

Males

Age	N	Mean Score	T-Value (with previous age)
12	182	558.51(S=147.25)	-----
13	144	780.66(S=151.44)	-----
14,15	30	739.67(S=169.19)	1.23

Table 75

Age Comparison for the Shuttle Run (September)

Females			
Age	N	Mean Score	T-Value (with previous age)
11	61	13.76(S=1.26)	-----
12	208	13.50(S=1.20)	1.28
13	159	13.45(S=1.61)	0.30
14,15	18	13.52(S=1.98)	-0.17
Males			
Age	N	Mean Score	T-Value (with previous age)
11	54	12.99(S=1.21)	-----
12	188	12.90(S=1.43)	0.45
13	162	12.33(S=1.20)	3.77**
14,15	35	12.23(S=1.60)	0.45

$^{**}p < .01$.

Table 76

Age Comparison for the Shuttle Run (May)

Females			
Age	N	Mean Score	T-Value (with previous age)
12	181	13.10(S=1.20)	-----
13	130	12.90(S=1.61)	1.25
14,15	11	13.39(S=1.98)	-0.81
Males			
Age	N	Mean Score	T-Value (with previous age)
12	182	12.40(S=1.43)	-----
13	144	12.09(S=1.20)	2.11*
14,15	30	12.00(S=1.60)	0.31

*$p < .05$.

Table 77

Fitness Change and Initial Fitness Level

	Overall			
	Above 50th Perc.		Below 50th Perc.	
	f_o	f_e	f_o	f_e
Improvement	276	286.12	319	308.88
No Improvement	101	90.88	88	98.12

Chi-Square = 2.91.

	Females			
Improvement	120	125.85	140	134.15
No Improvement	62	56.15	54	59.85

Chi-Square = 1.71.

	Males			
Improvement	156	160.11	179	174.89
No Improvement	39	34.89	34	38.11

Chi-Square = 1.13.

the teacher (Table 78). It was found that for those students above the 50th percentile, there was no significant relationship between fitness change and teacher qualifications. However, there was a significant relationship between fitness change and teacher qualifications for students below the 50th percentile.

Teacher Specialization and Fitness Change

One of the independent variables in this research was teacher specialization. Table 79 shows that there was a significant relationship (.01 level) between this variable and fitness change. As previously mentioned, qualified instruction was most significantly related to fitness change for those students who were below the 50th percentile in September.

When gender was included, the results varied. There was a significant relationship between teacher specialization and fitness change for females (.005 level), but not for males.

Instruction Time and Fitness Change

During the test period of this research, the average Norfolk student received 110 minutes of physical education instruction per six-day cycle. In order to assess the effect of instruction time on fitness change, two categories were compared: students receiving less than the average amount, and those receiving at least the average amount.

Table 80 shows that there was a significant difference (.005 level) between students receiving less than 110 minutes of

Table 78

Fitness Change and Teacher Specialization (Controlling for Initial Fitness Level)

Above 50th Percentile

	Specialized		Non-Specialized	
	f_o	f_e	f_o	f_e
Improved	211	207.80	29	32.20
Not Improved	73	76.20	15	11.80

Chi-Square = 1.37.

Below 50th Percentile

	Specialized		Non-Specialized	
	f_o	f_e	f_o	f_e
Improved	277	267.03	78	87.97
Not Improved	66	75.97	35	25.03

Chi-Square = 6.78, df = 1, $p < .01$.

Table 79

Fitness Change and Teacher Specialization

	Overall			
	Specialized		Not Specialized	
	f_o	f_e	f_o	f_e
Improvement	489	476.00	106	119.00
No Improvement	139	152.00	51	38.00

Chi-Square = 7.34, df = 1, $p < .01$.

	Females			
Improvement	215	203.69	46	57.31
No Improvement	80	91.31	37	25.69

Chi-Square = 9.24, df = 1, $p < .005$.

	Males			
Improvement	274	273.27	60	60.73
No Improvement	59	59.73	14	13.27

Chi-Square = 0.06.

Table 80

Fitness Change and Instruction Time

Overall

	Over 110		Under 110	
	f_o	f_e	f_o	f_e
Improvement	239	256.60	356	338.40
No Improvement	100	82.40	91	108.60

Chi-Square = 8.74, df = 1, $p < .005$.

Females

Improvement	114	119.10	145	139.90
No Improvement	58	52.90	57	62.10

Chi-Square = 1.32.

Males

Improvement	125	135.60	210	199.40
No Improvement	41	30.40	34	44.60

Chi-Square Value = 7.61, df = 1, $p < .01$.

instruction and those receiving more. This difference was due to the better performance by those receiving less instruction time than was expected.

Within this variable there were gender differences. Fitness change was related to instruction time for males (.01 level), but not for females.

Specialized instruction was an important variable with respect to instruction time. For those students receiving more than 110 minutes of instruction time, there was no relationship between fitness change and specialized instruction. However, for those students receiving less than 110 minutes, there was a significant relationship between fitness change and teacher qualifications (Table 81).

Teaching Experience and Fitness Change

The average physical education teacher in Norfolk had 18 years of teaching experience at the beginning of this research. In order to evaluate the effect of teaching experience on fitness change, two categories were established: teachers with less than 18 years of experience, and those with 18 or more years.

Table 82 indicates that there was a significant relationship (.005 level) between teaching experience and fitness change; students whose teachers had less than 18 years experience performed significantly better than those whose teachers had over 18 years of experience.

Neither gender showed a significant relationship between the

Table 81

Fitness Change and Teacher Specialization (Controlling for Instruction Time)

110 Minutes and Over

	Specialized		Non-Specialized	
	f_o	f_e	f_o	f_e
Improved	281	277.78	75	78.22
Not Improved	67	70.22	23	19.78

Chi-Square = 0.84.

Under 110 Minutes

	Specialized		Non-Specialized	
	f_o	f_e	f_o	f_e
Improved	207	197.28	32	41.72
Not Improved	72	81.72	27	17.28

Chi-Square = 9.37, df = 1, $p < .005$.

Table 82

<u>Fitness Change and Teaching Experience</u>

	Overall			
	Over 18 Years		Under 18 Years	
	f_o	f_e	f_o	f_e
Improvement	250	269.50	345	325.50
No Improvement	106	86.50	85	104.50

Chi-Square = 10.62, df = 1, $p < .005$.

	Females			
Improvement	134	142.70	125	116.30
No Improvement	72	63.30	43	51.70

Chi-Square = 3.84, df = 1, $p < .05$.

	Males			
Improvement	115	121.70	220	213.30
No Improvement	34	27.30	41	47.70

Chi-Square Value = 3.16.

two variables on its own, although both showed a better performance by students whose teachers had less than 18 years experience.

For students whose teachers had less than 18 years of experience, there was no significant relationship between fitness change and teacher qualifications. However, a significant relationship did exist between those two variables for students whose teachers had over 18 years of experience (Table 83).

Nature of the Class and Fitness Change

In Norfolk, 479 students were instructed in gender-segregated classes, while 306 participated in co-educational classes. As indicated on Table 84, there was no relationship for males or females between fitness change and the nature of the class.

In co-educational classes, there was no significant relationship between fitness change and the qualifications of the teacher (Table 85). In gender-segregated classes, however, there was a significant relationship between fitness change and teacher qualifications. Those students in segregated classes performed significantly better when the teacher had specialized physical education training.

Summary of Results

From the September results it can be stated that, according to Canada Fitness Award standards, Norfolk students were

Table 83

Fitness Change and Teacher Specialization (Controlling for Teaching Experience)

18 Years or Over

	Specialized		Non-Specialized	
	f_o	f_e	f_o	f_e
Improved	185	175.35	65	74.65
Not Improved	64	73.65	41	31.35

Chi-Square = 6.01, df = 1, $p < .05$.

Under 18 Years

	Specialized		Non-Specialized	
	f_o	f_e	f_o	f_e
Improved	303	303.99	42	41.01
Not Improved	75	74.01	9	9.99

Chi-Square = 0.13.

Table 84

Fitness Change and Nature of the Class

	Co-educational		Segregated	
	f_o	f_e	f_o	f_e
Overall				
Improved	238	231.94	357	363.06
Not Improved	68	74.06	122	115.94

Chi-Square = 1.08.

Females				
Improved	96	91.83	165	169.17
Not Improved	37	41.17	80	75.83

Chi-Square = 0.94.

Males				
Improved	142	141.97	192	192.03
Not Improved	31	31.03	42	41.97

Chi-Square = 0.00.

Table 85

Fitness Change and Teacher Specialization (Controlling for Nature of the Class)

	Gender-Segregated			
	Specialized		Non-Specialized	
	f_o	f_e	f_o	f_e
Improved	332	321.23	25	35.77
Not Improved	99	109.77	23	12.23

Chi-Square = 14.14, df = 1, $p < .001$.

	Co-educational			
	Specialized		Non-Specialized	
	f_o	f_e	f_o	f_e
Improved	156	152.94	82	85.06
Not Improved	40	43.06	27	23.94

Chi-Square = 0.78.

physically unfit; by virtually every categorization, they scored below standard. In May, the results indicated that these same students were fit.

There was no relationship between gender and total fitness in either September or May. In September, both genders were below CFA standards and in May, both were above. This represented significant improvements for both genders, but more so for males.

There was no relationship between total fitness and age in either September or May. However, in September, when specific CFA subtests were considered, these two variables were related in some cases. For males, there was no relationship between age and performance on any of the subtests, while, for females, there was a significant relationship between age and almost all CFA test scores. In many instances the significance of age with respect to performance was attributable to the substandard performance of thirteen year olds, the least fit group within Norfolk in September. In May the genders were more similar; for females the relationship became less definitive while, for males, the relationship became stronger.

The scores for each age group significantly improved over the course of the year.

Fitness and school showed a significant relationship in both September and May at the .001 level. Also, there was a significant relationship between school and fitness change.

Fitness and participation in extracurricular physical activity showed a significant relationship, especially for males. Extracurricular activity was also significantly related to fitness change.

In September, boys outperformed girls at each age level, in many instances significantly so. This pattern continued in May.

For boys, performance in September and May generally improved by age. With few exceptions each successive age outscored the previous one on all six tests. For girls, there was no definitive pattern in relation to age.

There was no relationship between grade and fitness in September or May. In September, both grades were significantly below standard, while, in May, both were above standard. The fitness change was significant for both grades.

There was no relationship between fitness change and the initial fitness levels of the students, unless teacher specialization was considered. For students below the 50th percentile in September, there was a significant relationship between fitness change and teacher specialization.

Fitness change and teacher specialization were significantly related for females and overall, but not for males.

Instruction time and fitness change were significantly related overall and for males, but not for females.

Teaching experience and fitness change were significantly related: students whose teachers had less than 18 years of experience improved more than the rest. In addition, when only teachers with over 18 years of experience were examined, there was a significant relationship between fitness change and teacher specialization.

There was no relationship between fitness change and nature of the class, for either males or females. However, when only gender-segregated classes were examined, there was a significant relationship between fitness change and teacher specialization.

CHAPTER V

CONCLUSIONS, RECOMMENDATIONS, AND SUMMARY

This chapter is a discussion of the Chapter IV findings, both expected and unexpected. A major focus is the comparison of these findings with those of other researchers. In some cases these results have educational implications for individual schools, as well as for the entire Norfolk system.

Youth Fitness

There is an apparent difference of opinion among researchers as to the fitness of students. Fitness Canada establishes standards based upon the results of the participants tested since the inception of the Canada Fitness Award; the last revision (1985) saw raised standards. This indicated that, firstly, since the previous revision, test scores had improved. Secondly, Canadian youth are improving in fitness. Gauthier, Massicotte, Hermiston, and MacNab (1983) concur with Fitness Canada on that point, as do the findings of this research.

Other researchers, conversely, have noted an overall decline in student fitness levels. The U.S. National Youth Fitness Study (1985), Murphy (1986), Cooney (1987), Hoerr (1987), and Giel (1988) have refuted the notion that student fitness levels are improving.

This fitness issue is clouded. While in Canada, there are differing opinions concerning this question--Fitness Canada (1985) for improvement; Cooney (1987) for decline--in the United States the verdict is clear: children are becoming less fit

(Giel, 1988; Hoerr, 1987; Murphy, 1986). Are there cultural factors which account for this nationalistic difference? It is possible that American youth are simply less fit than Canadians, but this finding has never been established. It is also possible that with increasing urbanization U.S. youth are becoming less active than Canadians; lower activity levels have been correlated with decreased fitness levels (Cooney, 1987; Gabbard et al., 1987; McGinnis, 1987; Rippe, 1987).

However, the most plausible explanation is that the two countries use different tests, thereby rendering meaningful comparisons virtually impossible, a dilemma acknowledged by Giel (1988). While it is true that different tests are employed--the most widely used test in Canada is the CAHPER Fitness-Performance Test, and in the U.S. the AAHPERD Health-Related Fitness Test is prevalent--it should be noted that the Canadian test was derived from the American one. The difference in results lies in the fact that American researchers, led by Pate and Ross, have moved toward health-related definitions of fitness, whereas Canadians still prefer the performance-related tests. While the difference in tests appears critical, Beauchamp (1980) stated that the difference between the two concepts is simply one of degree. The move away from performance testing is not supported by all researchers (Malina, 1988) but the most prominent among the AAHPERD researchers have moved that agency in a direction which has in turn focussed media attention upon the noteworthy issue of youth "unfitness". In short, then, there are not significant differences between American and Canadian youth fitness, only differences in testing instruments.

Fitness and Gender

It is clear from past research (Alexander, Ready, & Fougere-Mailey, 1985; Fitness Canada, 1985; Raithel, 1987) that males outperform females on virtually all tests of fitness, including the ones contained in this research. Chapter IV confirms these results. At this point, the issue with respect to fitness and gender, however, is not whether males outscored females in terms of raw scores, but rather whether males and females scored differently than expected when compared with Canada Fitness Award standards, which already take into account male-versus-females performance differences.

In Norfolk, gender was not a determinant of fitness in either September or May; there was no significant difference in the performance of the genders, when compared to Canada Fitness Award standards. Therefore, the null hypothesis--that there would be no significant difference between fitness and gender--must be accepted.

In September both genders were unfit, while in May both were fit. This change in fitness levels indicated that Norfolk did, in fact, have an effective physical education program for the school year 1988-89. (Therefore, the null hypothesis--that there would be no significant difference between pre- and post-treatment scores--is rejected.) This can be stated because the criterion of effectiveness chosen for this research was a significant improvement in fitness levels, represented not by significant individual improvement, but rather, improvement in a significant number of students. This choice was consistent with the Ministry of Education's (1978) criterion for effectiveness: improvement in the majority of students. The criterion of effectiveness for this research was simply more stringent.

Thirteen-year-old females were the least fit group in Norfolk in both September and May. It could be suggested that this particular group of girls just happened to be less fit than all other age groups. However, the thirteen-year-old female groups in September and May were not the same students. Approximately one half of the May thirteen year olds had been twelve in September. This being the case, it can be stated with certainty that thirteen-year-old females were the least fit group in Norfolk, regardless of the time of year. This finding coincides with existing literature: Bozzo (1983) and Cooney (1987) have both stated that shortly after puberty the fitness levels of females drop sharply.

The crux of the issue in this research was whether the physical education program was effective. There was a significant improvement in fitness levels for both genders, but more so for males. As stated earlier in this paper, one criterion by which the Ministry of Education deems a physical education program effective is the improvement in fitness levels of a majority of students. The criterion of effectiveness for this research, however, was an improvement in the fitness levels of a significant number of students.

It is clear that the physical education program in Norfolk was effective in 1988-89. By anyone's standards--Ministry of Education (1978), Pifer (1987), or this author--this cannot be refuted. However, these results could be slightly exaggerated.

It has been established that over the summer months students fail to retain much of the academic material presented throughout the year; during the early part of the following year, these students regain the level previously held. The same could hold true for physical fitness levels. By the beginning of the 1988-89 school year, the students could have lost some of the physical fitness achieved during the previous school year. Some students had participated in no physical activity during the summer, while others participated in activities which did not maintain the previous fitness levels.

It is also possible that the physical education instruction received during the previous year had been ineffective in developing the true fitness potential of the students.

A third possibility is that the physical education instruction received during the year regained the previous levels, and added to those levels. It is most likely that, while the physical education instruction was indeed effective, it could be slightly less so than shown in the results of this research. But, let there be no doubt: the program was effective.

Previous research (Corbin, 1980; Fitness Canada, 1985;

Raithel, 1987) has indicated that male fitness levels are generally higher than those of females. There has been little research, however, that discusses how fitness change relates to gender. From this research, it can definitely be stated that, not only were male fitness levels higher than female at the outset, they also improved more.

When individual Canada Fitness Award subtests are considered, it can be seen that in both September and May performances in the standing long jump and curl-ups were significantly related to gender. In May the scores were even more significant than in September.

These two subtests measure the explosive power of the legs (standing long jump) and the muscular endurance of the abdominals (curl-ups).

Of the performances on individual CFA subtests, all follow the same pattern - substandard performance in September which is much improved in May - except one: curl-ups. In September, males and females were significantly better than expected and there was actually an overall decrease in performance in May (due to a decrease by females). A possible explanation for this situation can be found in the fact that the curl-up subtest was a new item in the 1985 revised Canada Fitness Award battery of tests, consequently the number of students having taken this subtest was much lower than the others. This being the case, the norms established for the subtest might have been too low. This would account for the higher than expected scores in September.

Ian Craigon (1989), of the Canadian Association for Health, Physical Education, and Recreation (CAHPER), acknowledged that when the Canada Fitness Award is next revised, the curl-up subtest will receive much attention.

On the standing long jump there was a significant improvement overall, which was attributable to a similar improvement by males; there was no significant change in female performance. Therefore, the null hypothesis--that there would be no significant difference between pre- and post-treatment scores--is rejected. This would suggest that the Norfolk physical education program was far more beneficial to males than females in developing explosive leg power. Males have a greater muscle mass than females (Raithel, 1987); if the muscle mass of each gender is improved commensurately, it makes sense for males to improve more in absolute terms. It could be expected that male scores should be higher (Canada Fitness Award, 1985), but this research indicated that males also improved more.

On the 50-meter dash there was a significant improvement for both genders, but more so for males. Therefore, the null hypothesis--that there would be no significant difference between pre- and post-treatment scores--is rejected. The physical education program is more beneficial for males than females in improving speed. The previous argument concerning greater muscle mass in males (Raithel, 1987) helps explain the difference in improvement in 50-meter dash scores, since that event requires explosive leg power.

In push ups, there was a significant improvement for both males and females, but more so for males. Therefore, the null hypothesis--that there would be no significant difference between pre- and post-treatment scores--is rejected. This indicates that the physical education program was beneficial in developing arm strength and muscular endurance for both genders, but it was more beneficial for males. The previous argument concerning muscle mass (Raithel, 1987) again helps explain the difference in improvement.

On the endurance run, there was a significant improvement for both males and females, but more so for males. Therefore, the null hypothesis--that there would be no significant difference between pre- and post-treatment scores--is rejected. This indicates that the physical education program was beneficial in improving cardiorespiratory endurance, but it was more beneficial to males than females. Males have larger hearts than do females (Raithel, 1987); the result should be a generally more efficient cardiorespiratory system. If both males and females are exposed to cardiorespiratory training, it could be argued that the better system should improve more.

On the shuttle run, there was a significant improvement for both males and females, but more so for females. Therefore, the null hypothesis--that there would be no significant difference between pre- and post-treatment scores--is rejected. This indicates that the physical education program was beneficial in improving agility and it was more beneficial for females. Males experience their peak growth rate in height and body mass between ages twelve and thirteen (Shephard, 1986); these ages correspond closely to the ages of grades seven and eight boys. Females have already experienced this growth spurt by grade seven. The shuttle run is based upon agility (Fitness Canada, 1985), a fitness component affected by coordination, which can be greatly influenced by sudden growth spurts.

Fitness and Age

It had initially been a concern of this author that students at various stages of the most drastic growth spurts of their lives (Shephard, 1986) would function quite differently on tests of fitness. Evidence did not bear out this hypothesis.

When compared to Canada Fitness Award standards, there was no relationship between age and fitness. Therefore, the null hypothesis--that there would be no significant relationship between fitness and age--is accepted. The physical education program was generally equally effective for all age groups; in September, each age group was below standard, and in May each age group was above standard. This represented a significant improvement for each age group.

With no relationship between age and fitness, or fitness improvement, age need not be a concern when assigning students to physical education classes.

Fitness and Grade

Grades seven and eight students differ in terms of physical education experience. Grade seven students have largely come from a junior division where most physical education is taught by homeroom teachers, often untrained in the specifics of physical education. Grade eight students, on the other hand, have generally had one year of more intensive physical education instruction.

It could be suggested that the grade seven students should show more improvement because the fitness potential has been attained to a lesser degree.

The results, however, did not substantiate this hypothesis. There was no significant relationship between grade and fitness, nor grade and fitness change. Therefore, the null hypothesis--that there would be no significant relationship between grade and fitness--is accepted. This being the case, there should be no concern when combining grades seven and eight students in the same physical education class, a practice followed often in Norfolk.

There was a significant improvement in fitness for both female and male grade seven students; this indicates that the physical education program was effective for grade seven students. The same holds true for grade eight students.

Fitness and School

The school is an amalgam of so many factors (extracurricular physical activity, experienced instruction, instruction time, initial fitness level, teaching experience, and coeducational classes), many of which have documented effects on fitness change, that to expect all of these variables to behave similarly in each school is unreasonable. Consequently, a significant difference among schools was expected. This fact was substantiated by the Chapter IV results.

There was a significant relationship between fitness and school in September, and an even more significant relationship in May. Therefore, the null hypothesis--that there would be no significant relationship between fitness and school--is rejected. This indicates that the school which a student attended was a significant factor in determining the fitness, and fitness improvement, of that student.

More interestingly, there was not a significant relationship between the fitness of either gender and school in September, but there was in May, at the .001 level of confidence. This suggests that the physical education instruction at various schools widened the range of fitness levels within Norfolk. This being the case, it is obvious that the physical education instruction varied significantly in effectiveness from school to school.

While eight schools had programs which significantly improved fitness levels, it is evident that even among these schools some stood out in terms of effectiveness. Schools 4, 5, 6, and 11 were far above the others, with School 5 proving to be the most effective with respect to fitness improvement for all students.

When males were considered alone, eight schools provided instruction which significantly improved fitness levels. School 5 was the most effective school in improving the fitness levels of boys. For females, there were six schools which significantly improved the fitness levels of students, with School 6 being the most effective school.

Fitness and Extracurricular Physical Activity

Past research leads one to expect that those students who participated in extracurricular physical activity would score higher on fitness tests (Cooney, 1987; Gabbard et al., 1987; McGinnis, 1987; Rippe, 1987). That fact was borne out by this research. Another fact established was that those involved in extracurricular physical activity improved their fitness levels more than those not involved.

In both September and May there was a significant relationship between fitness and participation in extracurricular physical activity. Therefore, the null hypothesis--that there would be significant relationship between fitness and participation in extracurricular physical activity--is rejected. This does not necessarily indicate a causal relationship between the two variables, but a link between the two is irrefutable.

In September, the fitness of only one female age group was related positively with extracurricular physical activity, while for males only one group did not show a positive relationship.

In May, all but one group showed this positive relationship. It is felt by the author that this finding further adds to the evidence of an effective physical education program. In September, even after a large percentage of the students had participated in summertime physical activities, Norfolk students proved to be unfit; in May, again after a large percentage of students had participated in extracurricular physical activities, the Norfolk students proved to be fit. Assuming that the effect of extracurricular activity was approximately equal, the improvement in fitness levels must be attributable to some factor other than the extracurricular activity; arguably, this factor could be the physical education instruction received.

<u>Raw Scores (by Gender)</u>

According to past research, it was to be expected that males would prove physically superior to females on the Canada Fitness Award subtests (Alexander, Ready, & Fougere-Mailey, 1985; Raithel, 1987). (With the exception of flexibility, according to Raithel, 1987, and others, males generally outperform females on all fitness tests.) That fact was substantiated by this research.

Twelve-year-old boys were physically superior to their female counterparts, according to the criteria used in this research. Boys outscored girls on every Canada Fitness Award subtest during both of the testing periods. In addition, the gap between performances on all six subtests widened from September to May. Physical education instruction was more effective for twelve-year-old boys than girls.

Thirteen-year-old boys were physically superior to thirteen-year-old girls. Boys outscored girls on each of the six CFA subtests during both of the testing periods. Also, the gap between performances widened on four of the six subtests.

Fourteen-year-old boys were also physically superior to their female counterparts. While surpassing females on all six subtests during both testing periods, the gap widened on half of the tests over the course of the year.

As mentioned earlier, researchers have found that, with the exception of flexibility, males consistently outperform females on fitness tests. This can safely lead to the conclusion that, with respect to fitness tests, males are physically superior to females. Confounding this physiological reality, however, is the fact that males still have access to more sports programs (Raithel, 1987; Ross & Pate, 1987). This fact might help explain differences in fitness change between genders. A second factor which has been established, by researchers such as Corbin (1980), as affecting female performance is motivation. It sometimes appears that females could perform better if they really wanted to.

Raw Scores (by Age)

Researchers have shown that just after puberty fitness scores for males rise sharply (Bozzo, 1983; Greenberg & Pargman, 1986).

It is apparent from this research that there is a relatively predictable progression of performance levels for males. In the vast majority of cases, fourteen year olds outperform thirteen year olds, who in turn outperform twelve year olds, and so on. If one examines the Canada Fitness Award (1985) standards, these results were to be expected. This pattern of physical superiority by successive ages continues at least until age seventeen, according to Fitness Canada (1985); the intermediate school grades provide excellent grounds for comparison since all students fall within the age range where this pattern is constant. In contrast, Cooney (1987), in quoting the Saskatchewan Growth Study, cited a significant decline in fitness levels after the age of twelve.

Past research also has shown that, for females, performance improves with age until puberty; shortly thereafter, scores drop rapidly (Bozzo, 1983; Cooney, 1987; Corbin, 1980; Fitness Canada, 1985). This research also suggests that for females there is not the same consistency of age superiority as is true for males. One fact that helps account for the lack of a consistent pattern is the performance of the fourteen and fifteen year olds. Eleven year olds were only included in the September testing, but in all of the six subtests they were bettered by the twelve year olds. Thirteen year olds bettered twelve year olds on 80% of the comparisons. However, thirteen and fourteen year olds each outscored the other on 50% of the comparisons.

There is an educational consequence of this last finding: in this research it was found to be more important for females to have qualified instruction than for males. It can be suggested that pubescent females require qualified physical education teachers who are aware not only of the fact of decreasing female performance, but also possible strategies for compensating this effect. In a school where classes are segregated by grade then, it would be most sensible to have a qualified teacher with the grade eight girls.

Initial Fitness Level and Fitness Change

A point of concern for this author had been the effect of physical education instruction on students of varying abilities. While several researchers have alluded to the effects of genetics on fitness (Blair, Falls, & Pate, 1983; Cowart, 1987; Murphy, 1986; Shephard, 1986), how does fitness change relate to initial fitness level? In addition, physical education is a subject area which generates extremes in attitude; few students seem to be ambivalent about physical education.

With this in mind, it could be predicted that those students who enjoy physical education, and are proficient, would work harder, and thereby improve their fitness levels more. Conversely, it has been cited by several researchers (Cuniff, 1985; Pifer, 1987; Shephard, 1986) that less fit students have a lower threshold for improvement; thus, it should be easier for less fit students to show improvement than for more fit students.

This research partially substantiated the latter supposition; while the relationship was not significant, students whose initial fitness levels were below the 50th percentile did improve more than those above the 50th percentile. However, the null hypothesis--that there would be no significant relationship between fitness change and initial fitness level--is accepted.

In addition to the fact that less fit students have more room for improvement, it could be suggested that, as in classroom subjects, teachers focus upon weaker physical education students and encourage them to more improved scores.

The negative side of such a finding is that the results might also be suggesting that physical education classes have the effect of eliminating the extremes in fitness; the weaker students are brought more in line with the mainstream of students, while the students whose initial fitness levels are high are also brought closer to the average students because the instruction given is aimed at improving the weakest students, not challenging the better students. The existence of enrichment programs in all subject areas affirms the belief that teachers often ignore the students who can handle the work easily, thereby requiring an "enriched" setting where these students can also improve.

To further examine the issue of initial fitness level, student scores were analyzed, while controlling for teacher specialization. For students above the 50th percentile in September, there was no significant relationship between fitness change and teacher specialization; for students below the 50th percentile, there was such a relationship. This suggests that for students who are already physically fit, the qualifications of the teacher are of little importance. However, it is critical that unfit students receive qualified instruction.

Teacher Specialization and Fitness Change

As was expected, based upon past research (Cooney, 1987; Hansen, 1990; Pate & Ross, 1987), students in Norfolk scored better in terms of fitness improvement when they were taught by teachers who were qualified in the field of physical education. Therefore, the null hypothesis--that there would be no significant relationship between fitness change and teacher specialization--is rejected. This significant relationship was the result of the positive relationship for females. For males, there was no relationship between teacher specialization and fitness improvement. In other words, it made no difference to males whether the teacher was qualified or not. For females, it was imperative that the teacher be qualified.

This would suggest that where physical education classes are segregated by gender, the female classes should always have qualified instruction.

Instruction Time and Fitness Change

Since the classic Vanves (1951) experiment into the effects of daily physical education instruction, there has been a controversy concerning the appropriate amount of physical education instruction time. Increasing the amount of instruction time would only make sense if, in fact, it could shown that such an increase were beneficial.

In Norfolk, the average instruction time was 110 minutes per six-day cycle. Those students who received more than the average amount did not improve as much as did those who received less than the average amount. However, the null hypothesis --that there would be no significant relationship between fitness and instruction time--is accepted. Since this result was unexpected, the results were further analyzed. Controlling for instruction time, fitness change and teacher specialization were examined. It was discovered that instruction time is, in fact, an important variable: for students receiving less than 110 minutes of instruction, it was essential that the teacher be specialized in physical education.

Considering this point, and the fact that several physical education teachers in Norfolk are not specialized, it is clear that at the minimum all schools should provide at least 110 minutes of instruction time, thereby providing equal opportunity for fitness improvement. More importantly, considering that physical education instruction in Norfolk is effective, there should be an overall move toward daily physical education instruction.

The move toward daily physical education has received much publicity recently (Hansen, 1990; McGinnis, 1987; Murphy, 1987;

OPHEA, 1988; Ross & Pate, 1987), but in Norfolk this would represent a drastic change. Assuming an average period length of 40 minutes, Norfolk students received in 1988-89 less than one period of physical education every other day, approximately the Canadian average over the last ten years, according to Beauchamp (1980), and insufficient to improve physical fitness, according to the Ontario Medical Association (1987); daily physical education would require 240 minutes per cycle. It might be more realistic to campaign for 120 minutes per cycle one year, analyze the results, then, if warranted, push for an additional period per cycle. In this manner, the move toward daily physical education would be gradual and sensible. Practically speaking, no administrator would, or should, agree to allot daily physical education time without proof of its effectiveness. This research, however, certainly supports the notion that "more is better" in Norfolk.

The wide range of instruction time offered to Norfolk students--a minimum of 80 minutes per cycle through a maximum of 160 minutes per cycle--reflects the perceived importance of physical education by the particular school administrator. Any campaign to increase physical education instruction time should be aimed at upper administration, where the power exists to mandate and nurture system-wide increases. In this manner students need not be penalized by virtue of being situated in a school whose administrator holds physical education in low esteem. Pate and Ross (1987) and Hansen (1990) furthered this argument by suggesting that fitness is affected by the investment of resources (time), or lack thereof.

Interestingly, instruction time was not a significant factor for females. As mentioned earlier, males improved their fitness levels regardless of the qualifications of the teacher. Males benefit from the increased instruction simply by having the opportunity to participate more; for females, increased time was a factor with a qualified teacher, but not without.

Teaching Experience and Fitness Change

It would be logical to assume that most teachers would be more effective as their experience grew. While this is undoubtedly true, there exists a point beyond which a physical education teacher's effectiveness can decline, insofar as fitness improvement of students is concerned.

For this research the average Norfolk physical education teacher had 18 years of teaching experience. Students whose teachers had less than 18 years of experience improved significantly more than those whose teachers had more than 18 years. Therefore, the null hypothesis--that there would be no significant relationship between fitness and teaching experience--is rejected. It could be justly proposed that younger teachers are more likely to be qualified physical education instructors; the significant difference, however, surpasses the effect of that factor.

Physical education teachers spend more time involved in extracurricular activities than do most other teachers (Grant, 1990). In a sense, a physical education teacher with 18 years of experience has spent far more time in school than the average teacher with the same years of experience; this can cause a type of "burn out".

We could all cite examples of teachers whose energy levels are chronically low. Hopefully, these teachers would also recognize this change and do as a 15-year veteran did when this author was in his first year: he changed subject areas. When asked if he had lost his love of athletics, he vehemently responded to the negative; he acknowledged though that he had "lost the spunk to do the job right". Physical education teaching is a job which demands a high level of energy and enthusiasm, in order to lead by example and to motivate the students to their best performances.

It behooves administrators to recognize attitudinal change and act accordingly. This is not to say that upon reaching an arbitrary age or experience level, a teacher should be forced to leave his/her preferred field. Nor does it mean that this should be the consequence after seeing an older teacher have a low-energy day. However, if an administrator witnesses a constant condition such as previously described, he/she must act in the best interest of the students. Offer experienced teachers new challenges: an assignment in physical education at another school, or a different subject area at the same school.

Nature of the Class and Fitness Change

While little literature was found on the subject, it had been a belief of this author that students who received physical education instruction in a coeducational setting would improve less in fitness than students in a gender-segregated setting. This notion was supported by the Ontario Medical Association (1987). However, that was not the case in this research: there was no difference between the fitness change in coeducational students and that of other students. Therefore, the null hypothesis--that there would be no significant relationship between fitness change and nature of the class--is accepted.

This suggests that administrators need not be concerned generally with having males and females in the same class, even though males and female fitness levels vary; each gender seems to approach the subject at its own level.

The issue of coeducational instruction was further examined in relation to teacher specialization. It was shown that for coeducational classes, there was no relationship between fitness change and teacher specialization. However, for segregated classes this relationship did exist.

These findings indicate that in coeducational classes the qualifications of the teacher are not significantly related to the fitness improvement of the student. In gender-segregated classes, the qualifications of the teacher were important: students whose teachers were specialized improved their fitness levels significantly more than did those students whose teachers were not specialized.

It was earlier stated that it was more important for females to have specialized instruction than for males. This being the case, it is suggested that much of the significance of this finding is due to the lack of fitness improvement by females who did not have a qualified physical education teacher. This further shows how critical it is that females be taught by qualified teachers. Again, males are not as greatly influenced by the qualifications of the teacher.

Suggestions for Future Related Research

Replication of this research

It would be useful to replicate this study, again in Norfolk and elsewhere, in order to confirm the results or prove they represent an anomaly. It might be that the students were in some way atypical of "normal" intermediate students; this might especially be warranted when one considers that the success of the Norfolk physical education program seems to refute research that indicates that physical education courses do little to alleviate the problem of declining fitness levels in youth (Beauchamp, 1980; Iowa State Department of Public Instruction, 1985).

It would also be interesting to replicate this study with younger students in order to develop a fitness profile of the average child as he/she moves from kindergarten through grade eight. In this manner it would be possible to locate grades where more expert instruction would be critical. Within most schools in Norfolk, primary and junior classroom teachers, with no physical education training, teach the class physical education. Hours of mis- or undirected physical education instruction might not be beneficial, especially for the "physically at-risk" student.

It would have been suggested that replication of the research would also have been warranted in the high school grades. However, it would be difficult to assess the effectiveness of a program with an overall participation rate of 10-25% (Guidance Departments of Simcoe Composite School, Delhi District Secondary School, Waterford District High School, Port Dover Composite School, and Valley Heights Secondary School, 1989). Even in light of lobbying by CAHPER (1988) and other physical education researchers (OPHEA, 1988) on the behalf of mandatory physical education instruction throughout all grades, the Ministry of Education indicates its position by mandating students to take only one physical education course throughout the high school years. The consequences of this stance can be appreciated when one considers that, for most teenagers, 80 to 100% of their physical activity comes from physical education courses (Bozzo, 1983; CDDS, 1983), largely because the adult fitness craze has not trickled down to children (Ross & Pate, 1987).

It should by noted that even in elementary school biases exist: there are established guidelines outlining the minimum number of periods per cycle for most subjects, but not physical education. It seems that physical education is sometimes treated as a "timetable filler". Hansen (1990) also acknowledged the "low status" of physical education in the school curriculum.

Academics and fitness

It would also be interesting to investigate the link between academics and fitness. Since the Vanves (1951) study over thirty years ago, this relationship has intrigued researchers, and the results have proved encouraging for physical educators: while causation has not been firmly established, nevertheless, there is a correlation between academic achievement and increased physical education instruction time (Gabbard et al., 1987; Murphy, 1987a). It would prove enlightening for skeptics to learn of the positive relationship between these two variables.

Physical education consultants

Cooney (1987) cited a decreasing number of consultants as one of the factors indicative of the declining state of physical education. It suggests the current esteem in which physical education is held. In Norfolk, there has never been a physical education consultant. Apparently this need has never been considered a priority.

Within a Board of Education where unqualified physical education teachers are operating, it would seem a necessity to have a qualified person offering assistance. The results of this research support the notion that inexperienced teachers could use the advice of an expert. With this finding in mind, it is the suggestion of this author that the Norfolk Board of Education create the position of Physical Education Consultant.

In addition to the assistance provided by a physical education consultant, it would be beneficial for physical education teachers whose programs are weak to be made aware of more successful ones. It would be advantageous to gain access to programs which are effective overall, but, in particular, it would be most useful to address individual components. If a program, for example, were weak in developing muscular endurance, the teacher involved could approach another teacher, whose program was shown to improve scores on the push up subtest, which is a measure of muscular strength and endurance.

Conclusion

This research was an attempt to evaluate the effectiveness of the Norfolk Board of Education physical education program, using fitness-level improvement as the criterion of effectiveness. This criterion is consistent with Ministry of Education (1978) policy, as well as with researchers who assert that physical education programs can be evaluated by measuring fitness (CDDS, 1983; Elson, 1981; Fox et al., 1987).

Developing physical fitness must be a concern for physical educators for another reason: fitness improves the quality of life (Falls, 1980; Grant, 1990). In the broadest sense, physical education, if properly presented, has the ability to benefit youngsters for life. Hinkle and Tuckman (1987) concur.

With the Canada Fitness Award as the measurement instrument, this research indicated that the Norfolk physical education program was, in fact, effective in terms of fitness improvement. The overall effectiveness of the program resulted from several variables: plentiful instruction time; qualified teachers; youthful teachers; and provision of extracurricular physical activity.

Even considering the fact that it is effective in improving fitness, the Norfolk physical education program is certainly not perfect. There are obvious weaknesses, largely centred upon educational inequities. All students should be presented an equal opportunity to improve their fitness levels; to this end, several considerations should be given:

Instruction should not be provided by older, inexperienced teachers; the consequences are well documented in this research.

Standardization of physical education instruction is needed; instruction time should not be left to the discretion of school administrators. As well, financial resources should not be allocated at the school level; system-wide allocation would neutralize any negative bias at the school level. In short, the Norfolk Board of Education must establish as a priority the development of a consistent physical education program, in terms of resources, instruction time, and teacher qualifications. If standardization of program is not the preferred option, then the Board should provide a consultant, whose positive effect would prove beneficial in limiting the consequences of inadequate instruction time, limited resources, and deficient teacher qualifications.

Finally, there are several criteria by which a physical education program can be evaluated; fitness improvement is but one. Other criteria of effectiveness were not addressed within this research. There is no way to speculate as to Norfolk's effectiveness with respect to these criteria. However, given the chosen criterion, it must be concluded that the Norfolk physical education teachers of grades seven and eight students are doing a creditable job.

REFERENCES

Books

Best, J. W., & Kahn, J. V. (1986). Research in education (5th ed.). Englewood Cliffs: Prentice-Hall.

Bosco, J. S., & Gustafson, W. F. (1983). Measurement and evaluation in physical education, fitness, and sports. Englewood Cliffs: Prentice-Hall.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research. Chicago: McNally.

Corbin, C. (1980). A textbook of motor development (2nd ed.). Dubuque: Brown.

Corbin, C. B., Dowell, L. J., Lindsey, R., & Tolson, H. (1981). Concepts in physical education. Dubuque: Wm. C. Brown.

Cuniff, D. E. (ed.) (1985). Implementation of health fitness exercise programs. AAHPERD.

Cureton, T. K., Jr. (ed.) (1983). Human performance: Efficiency and improvements in sports, exercise, and fitness. AAHPERD.

Devore, J., & Peck, R. (1986). Statistics: The exploration and analysis of data. St. Paul: West.

Elson, P. R. (1981). A simple guide to fitness. Toronto: Is Five.

Ferguson, G. A. (1971). Statistical analysis in psychology and education (3rd ed.). New York: McGraw-Hill.

Fox, E. L., Kirby, T. E., & Fox, A. R. (1987). Bases of fitness. New York: MacMillan.

Gabbard, C., LeBlanc, E., & Lowy, S. (1987). Physical education for children: Building the foundation. Englewood Cliffs: Prentice-Hall.

Greenberg, J. S., & Pargman, D. (1986). Physical fitness: A wellness approach. Englewood Cliffs: Prentice-Hall.

Guenther, W. C. (1964). Analysis of variance. Englewood Cliffs: Prentice-Hall.

Holland, G. J., & Davis, E. C. (1965). Values of physical activity (3rd ed.). Dubuque: Wm. C. Brown.

Hunsicker, P. (1963). Physical fitness. Washington: National Education Association.

Jewett, A. E. & Bain, L. (1985). The curriculum process in physical education. Dubuque: Wm. C. Brown.

Lazarus, M. (1982). Evaluating educational programs. Arlington: American Association of School Administrators.

Percival, J., Percival, L., & Taylor, J. (1977). The complete guide to total fitness. Scarborough: Prentice-Hall.

Pollock, M. L., Wilmore, J. H., & Fox, S. M. (1979). Health and fitness through physical activity. New York: John Wiley & Sons.

Sharkey, B. J. (1979). Physiology of fitness: Prescribing exercise for fitness, weight control, and health. Champaign: Human Kinetics.

Shephard, R. J. (1986). Fitness of a nation: Lessons from the Canada fitness survey. New York: Karger.

Thomas, J. R., & Nelson, J. K. (1985). Introduction to research in health, physical education, recreation, and dance. Champaign: Human Kinetics.

Journals

Alexander, M. J., Ready, A. E., & Fougere-Mailey, G. (1985). The fitness levels of females in various age groups. Journal of the Canadian Association for Health, Physical Education, and Recreation, March/April, 8-12.

Bailey, D., & Shephard, R. (1976). The Canadian home fitness test. Journal of the Canadian Association for Health, Physical Education, and Recreation, January/February, 20-21.

Blair, S. N., Falls, H. B., & Pate, R. R. (1983). A new physical fitness test. The Physician and Sportsmedicine, 4, 87-95.

Caspersen, C. J. (1987). Physical inactivity and coronary heart disease. The Physician and Sportsmedicine, 11, 43-44.

Cooney, D. (1987). Quality, daily physical education...the need is obvious. Journal of the Ontario Physical and Health Education Association, Winter, 24-25.

Cowart, V. S. (1987). How does heredity affect athletic performance? The Physician and Sportsmedicine, 4, 134-138.

Dahlgren, W. J. (1982). The Canada fitness award: Its development and use. Journal of the Canadian Association for Health, Physical Education, and Recreation, September/October, 3.

Docherty, D., & Quinney, A. (1984). Evaluating physical fitness with reference to the functional capacities of muscle. Journal of the Canadian Association for Health, Physical Education, and Recreation, March/April, 15-18.

Falls, H. B. (1980). Modern concepts of physical fitness. Journal of Physical Education and Recreation, April, 25-27.

Gauthier, R., Massicotte, D., Hermiston, R., & MacNab, R. (1983). The physical work capacity of Canadian children, aged 7 to 17, in 1983: A comparison with 1968. Journal of the Canadian Association for Health, Physical Education, and Recreation, September/October, 2-7.

Giel, D. (1988). Is there a crisis in youth fitness - or fatness? The Physician and Sportsmedicine, 10, 145.

Grace, N. (1987). O.M.A. position paper on the school physical education program. Journal of the Ontario Physical and Health Education Association, Spring, 13-16.

Grant, B. (1990). Influences on physical education and those who teach it. Journal of the Canadian Association for Health, Physical Education, and Recreation, March/April, 4-9.

Hansen, H. (1990). Barriers to QDPE and how to overcome them. Journal of the Canadian Association for Health, Physical Education, and Recreation, March/April, 16-21.

Hinkle, J. S. & Tuckman, B. W. (1987). Children's fitness: Managing a running program. Journal of Physical Education, Recreation, and Dance, January, 58-61.

Hoerr, S. M. (1987). Counselling obese adolescents. The Physician and Sportsmedicine, 7, 201-202.

Jette, M., Sidney, K., & Cicutti, N. (1984). A critical analysis of sit ups: A case for the partial curl-up as a test of abdominal muscular endurance. Journal of the Canadian Association for Health, Physical Education, and Recreation, September/October, 4-8.

Koslow, R. E. (1988). Can physical fitness be a primary objective in a balanced physical education program? Journal of Physical Education, Recreation, and Dance, April, 75-77.

Lacy, E., & Marshall, B. (1984). Fitnessgram: An answer to physical fitness improvement for school children. <u>Journal of Physical Education, Recreation, and Dance</u>, January, 18-19.

Lashuk, M. (1984). A percentile method of grading physical education. <u>Journal of the Canadian Association for Health, Physical Education, and Recreation</u>, March/April, 8-11.

Leger, L. (1984). The Canadian home fitness test revisited. <u>Journal of the Canadian Association for Health, Physical Education, and Recreation</u>, July/August, 6-8.

Londeree, B. R. (1981). Strength testing. <u>Journal of Physical Education, Recreation, and Dance</u>, <u>52</u>, 44-45.

Loughery, T. J. (1987). Evaluating program effectiveness. <u>Journal of Physical Education, Recreation, and Dance</u>, <u>58</u>, 63-64.

Malina, R. M. (1988). The average child deserves more fitness programs. <u>The Physician and Sportsmedicine</u>, <u>10</u>, 39.

Martens, F. L. (1982). Daily physical education - a boon to Canadian elementary schools. <u>Journal of Physical Education, Recreation, and Dance</u>, March, 55-58.

McGinnis, J. M. (1987). Summary of findings from the national children and youth fitness study II. <u>Journal of Physical Education, Recreation, and Dance</u>, November/December, 50.

Meyers, C. R. (1980). In quest of sound fitness testing. <u>Physical Educator</u>, <u>37</u>, 69-75.

Murphy, P. (1986). Youth fitness testing: A matter of health or performance? <u>The Physician and Sportsmedicine</u>, <u>5</u>, 189-190.

Murphy, P. (1987a). The Cincinnati Academy of Physical Education: Using fitness to aid learning. <u>The Physician and Sportsmedicine</u>, <u>4</u>, 173-176.

Murphy, P. (1987b). Fit to fight? <u>The Physician and Sportsmedicine</u>, <u>15</u>, 42.

Ontario Physical and Health Education Association (ed.)(1988). <u>Journal of the Ontario Physical and Health Education Association</u>, Fall, 9.

Pate, R. R. (1983). A new definition of youth fitness. <u>The Physician and Sportsmedicine</u>, <u>4</u>, 77-83.

Pate, R. R., & Ross, J. G. (1987). The national children and youth fitness study II: Factors associated with health-related fitness. Journal of Physical Education, Recreation, and Dance, November/December, 93-95.

Pifer, S. (1987). Secondary physical education - A new design. Journal of Physical Education, Recreation, and Dance, August, 50-51.

Plowman, S. A., Hastad, D. N., & Marett, J. R. (1983). A workshop model that works: Implementation strategies for health-related fitness testing. Journal of Physical Education and Recreation, February, 46-47.

Pritchard, J. (1987). Program planning for quality daily physical education. Journal of the Canadian Association for Health, Physical Education, and Recreation, November/December, 14-19.

Raithel, K. S. (1987). Are girls less fit than boys? The Physician and Sportsmedicine, 11, 156-157.

Raithel, K. S. (1988). Physical best: An assessment and educational program - AAPHERD issues a new fitness test. The Physician and Sportsmedicine, 16, 31.

Rippe, J. M. (1987). The health benefits of exercise (Part 2). The Physician and Sportsmedicine, 11, 120-125.

Ross, J. G. & Pate, R. R. (1987). The national children and youth fitness study II: A summary of findings. Journal of Physical education, Recreation, and Dance, November/December, 51-56.

Shephard, R. J., Volle, M., Lavallee, H., LaBarre, R., Jequier, J. C., & Rajic, M. (1982). Curricular time for physical education? Journal of Physical Education, Recreation, and Dance, November/December, 19-20, 28.

Smith, E. L., & Gilligan, C. (1987). Effects of inactivity and exercise on bone. The Physician and Sportsmedicine, 15, 91-92.

Steinhardt, M. A., & Stueck, P. M. (1986). Personal fitness: A curriculum model. Journal of Physical Education, Recreation, and Dance, September, 23-28.

West, B. H., Beveridge, S. K., & Workman, C. (1987). Color coding children's fitness performance. Physical Educator, 44, 400-404.

Research Papers/Documents

American Alliance for Health, Physical Education, and Recreation (1965). Youth fitness test manual. Washington, D. C.: Department of the National Education Association.

Beauchamp, L. S. (1980). Understanding fitness: A primer for school program planners. University of Alberta.

Bozzo, R. (1983). Evaluation of programs for the promotion of physical fitness and exercise. Washington, D. C.: Granville Corporation.

Committee for the Development of Sport (1983). Conclusions from work undertaken and texts approved by the CDDS. Strasbourg, France: Council of Europe.

Fitness Canada (1985). Canada fitness award: The inside story. Ottawa: Ministry of Supply and Services Canada.

Goode, R. C. (1976). Observations and suggestions on the physical fitness of our school children. A paper presented to the 53rd convention of the Canadian Education Association, Halifax, Nova Scotia, September 1976.

Griffin, T. B., Klinzing, J. E., & Ziegler, S. B. (1983). A comparative study on the effects of traditional vs holistic physical fitness programs. Cleveland State University.

Harrer, B. G. (1983). Factors influencing physical fitness evaluation. Indiana University.

Hastad, D. N., Marett, J., & Plowman, S. A. (1983). Evaluation of the health related physical fitness status of youth in the state of Illinois. DeKalb: Northern Illinois University.

Iowa State Department of Public Instruction (1985). A tool for assessing and designing a K-12 physical education program. Des Moines.

Ministry of Education, Ontario (1978). Curriculum guideline for the intermediate division: Physical and health education.

Ministry of Education, Ontario (1986). Ontario academic courses: Physical and health education (Validation Draft).

Ministry of Education, Ontario (1989). Ontario schools: Intermediate and senior divisions (grades 7-12/O.A.C.'s).

Tallmadge, G. K. (1977). The joint dissemination review panel ideabook. Washington: United States Government Printing Office.

Van Holst, A. (1983). Physical education curriculum for elementary grades. London, Ontario: University of Western Ontario.

Discussions

Craigon, I. (1989). Program Director, Quality Daily Physical Education, Canadian Association for Health, Physical Education, and Recreation.